LIMITING PATENTEES' MARKET POWER WITHOUT REDUCING INNOVATION INCENTIVES: THE PERVERSE BENEFITS OF UNCERTAINTY AND NON-INJUNCTIVE REMEDIES

*Ian Ayres**
and
*Paul Klemperer***

TABLE OF CONTENTS

INTRODUCTION 986
A. *The Intuition: Why It Is Possible to Restrict Patentees' Market Power Without Reducing Innovation Incentives* 989
B. *Implementation: How Probabilistic Patents Restrain Patentees' Market Power* 993
I. A SIMPLE MODEL 994
A. *Introducing Uncertainty Without Changing the Patent's Duration* 1000
B. *Lengthening the Patent's Life to Maintain the Innovation Incentive* 1001
C. *The Government Has Sufficient Information to Implement Probabilistic Patents* 1007
D. *Uncertainty Creates Delay* 1012
II. CAVEATS 1013
A. *Exacerbating Litigation Costs* 1014
B. *Probabilistic Patents Might Hinder Follow-On Innovation* 1014
C. *Infringers' High Costs May Dissipate Probabilistic Benefits* 1015

* William K. Townsend Professor, Yale Law School. B.A. 1981, J.D. 1986, Yale; Ph.D. (Economics) 1988, MIT. — Ed.

** Edgeworth Professor of Economics, Oxford University. B.A. 1978, Cambridge; M.B.A. 1982, Ph.D. (Economics) 1986, Stanford. — Ed. Many colleagues including Avery Katz, Ted Brown, Brian Coggio, Rebecca Eisenberg, Louis Kaplow, Rob Merges, Gideon Parchomovsky, Steve Salop, Daniel Sgroi, Lucy White, and seminar participants at Berkeley, Chicago-Kent, Georgetown, Michigan, NBER, and University of Miami provided helpful comments.

D. *Existing Deadweight Losses May Not Be Severe* 1016
III. IMPLICATIONS OF THE THEORY 1019
A. *Reducing the Use of Preliminary Injunctions* ... 1020
B. *Vesting Decisionmaking Authority in Less Specialized Hands* 1023
C. *Preferring Underinclusive Standards to Either Overinclusive Standards or Rules* 1024
D. *Allowing Patent Holders to Leverage Patent Market Power Into the Future* 1026
IV. ALTERNATIVE IMPLEMENTATIONS 1027
A. *Expanding Geographic or Product Scope Instead of Duration* 1027
B. *Partial Damages* 1028
C. *Duopoly Auction* 1031
CONCLUSION 1032

INTRODUCTION

Uncertainty and delay in patent litigation may have unforeseen virtues. The combination of these oft-criticized characteristics[1] might induce a limited amount of infringement that enhances social welfare without reducing (or without substantially reducing) the profitability of the patentee. Patent infringement is generally viewed as socially inefficient because infringement reduces the patentee's ex ante incentive to innovate. Limited amounts of infringement combined with increased patent duration, however, can substantially reduce the distortionary ex post effects of supra-

1. *See, e.g.*, Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698, 699 (1998) ("Long delays between the filing and issuance of biotechnology patents aggravate the problem of concurrent fragments. During this period of pendency, there is substantial uncertainty as to the scope of patent rights that will ultimately issue."). However, uncertainty has already been shown to be desirable in some special cases. First, Howard F. Chang, *Patent Scope, Antitrust Policy, and Cumulative Innovation*, 26 RAND J. ECON. 34 (1995); Jerry R. Green & Suzanne Scotchmer, *On the Division of Profit in Sequential Innovation*, 26 RAND J. ECON. 20 (1995); and Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSP. 29 (1991), have shown that uncertainty about whether follow-on innovation would infringe a patent might be desirable. *See infra* note 73. Second, uncertainty about which of two competing innovators' patent-claims on innovation will be valid is equivalent to each innovation receiving a small prize in expectation, which might attract more innovators to enter an R&D contest. *See* Richard Gilbert & Paul Klemperer, *An Equilibrium Theory of Rationing*, RAND J. ECON. (forthcoming 1999). Third, uncertainty about which claim is valid (which is economically equivalent to joint ownership of the rights to produce) can result in more efficient bargaining over the rights. *See* Ian Ayres & Eric Talley, *Solomonic Bargaining: Dividing a Legal Entitlement To Facilitate Coasean Trade*, 104 YALE L.J. 1027 (1995); Peter Cramton et al., *Dissolving a Partnership Efficiently*, 55 ECONOMETRICA 615 (1987).

competitive pricing without reducing the patentee's ex ante incentives to innovate. Indeed, this Article derives a legal regime that preserves the incentive to innovate by giving the patentee the same expected profits, but that substantially increases efficiency in comparison with an "idealized" patent regime (in which a patentee can instantaneously win an injunction to stop infringement).

Legal scholars have failed to appreciate that unconstrained monopoly pricing is not a cost-justified means of rewarding patentees.[2] The last bit of monopoly pricing produces large amounts of deadweight loss for a relatively small amount of patentee profit. If society wants to use patent profits to induce innovation, it should choose the method of producing a particular level of profit that produces the least cost to society. But allowing patentees to raise price all the way to the monopoly level is a little like giving them a license to steal car radios — it produces a social cost (to car owners) far greater than the private benefit. The dual thesis of this Article is that (1) efficient patent policy should strive to give patentees constrained market power, and (2) an enforcement regime with uncertainty and delay (in conjunction with the appropriate patent duration) may be one way of achieving this policy goal. Even if readers ultimately are unwilling to accept uncertainty and delay as tools to constrain patentees' market power, the insight that the last bit of monopoly pricing provides disproportionately small profits in comparison to its social cost should be of continuing independent concern to policymakers seeking an optimal enforcement regime.

Although this Article begins with a reductive mathematical model to show how uncertainty and delay can induce limited amounts of infringement, the purpose is not to propose a system in which the validity of patents is uncertain and only resolved ex post. Instead, appreciating how uncertainty and delay can constrain the patentee's market power can illuminate practical policy issues. For example, our model suggests ways to distinguish between the utility of different types of uncertainty. "Type I" uncertainty (which increases the chance that valid patents will not be enforced) is likely to be more efficient than "Type II" uncertainty (which increases the

2. Economists have understood that reducing the price of patented products with compulsory licensing and increasing the patent length (to maintain the patentee's expected profit) can increase welfare. *See, e.g.*, Richard Gilbert & Carl Shapiro, *Optimal Patent Length and Breadth*, 21 RAND J. ECON. 106 (1990); Paul Klemperer, *How Broad Should the Scope of Patent Protection Be?* 21 RAND J. ECON. 113 (1990); Pankaj Tandon, *Optimal Patents with Compulsory Licensing*, 90 J. POL. ECON. 470 (1982). The core insight of this Article is that delay and uncertainty can have the same beneficial effects on limiting the patentee's exploitation of market power.

chance that invalid patents will be enforced).[3] This suggests that recent efforts of the Federal Circuit to expand the use of preliminary injunctions may not be cost-justified — as they are likely to expand Type II uncertainty and reduce Type I uncertainty. The costs of Type I uncertainty and delay are not as great as previously thought, and we may not want to expend substantial resources to eliminate this feature of the current system. Appreciating the benefits of uncertainty argues against giving patentees a presumptive right to preliminary injunctions against infringement, and militates instead in favor of relatively stingy ex post measures of make-whole damages. Underinclusive "standards" for determining when a patent is valid are likely to be more efficient than either "rules" (which resolve uncertainty ex ante) or overinclusive standards.

Courts often argue that (1) patents, as a form of property, must include the right to injunctively exclude, and (2) failure to grant injunctions will lead to unlimited infringement, which would substantially undermine a patentee's incentive to innovate.[4] We contest both of these claims. Instead of accepting the essentialist notion that the "very nature" of property is the right to exclude, we suggest that the nature of patents should be to offer sufficient rewards to stimulate innovation. Denying immediate injunctive relief and substituting delayed probabilistic determination with monetary damages will only produce limited amounts of infringement and need not (especially when combined with extended patent life) reduce the ex ante incentive to innovate.

To highlight the perverse benefits of uncertainty and delay, this Article purposely abstracts from addressing which innovations are deserving of protection or what reward for innovation would be optimal.[5] Patent policy has to resolve two core issues: (1) how much of a reward should be granted to induce sufficient innovation; and (2) how can this desired reward be produced for the innovator with the least social inefficiency. This paper takes on the second question. Regardless of how much profit society decides to grant a pat-

3. The terms "Type I" and "Type II" are inspired by classical hypothesis testing. As traditionally defined, the possibility that a true null hypothesis will be rejected is referred to as a "Type I" error, while the possibility that a false null hypothesis will not be rejected is referred to as a "Type II" error. *See* James E. Krier, *Risk and Design*, 19 J. LEG. STUD. 781, 784 (1990); *infra* note 85. We are implicitly defining the null hypothesis to be that a patent is valid, but we have no strong view on which type of error should be defined as Type I or II.

4. *See infra* section III.A.

5. For leading analyses of these issues, see WILLIAM D. NORDHAUS, INVENTION, GROWTH, AND WELFARE (1969); F.M. Scherer, *Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation*, 62 AM. ECON. REV. 422 (1972); and Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839 (1990).

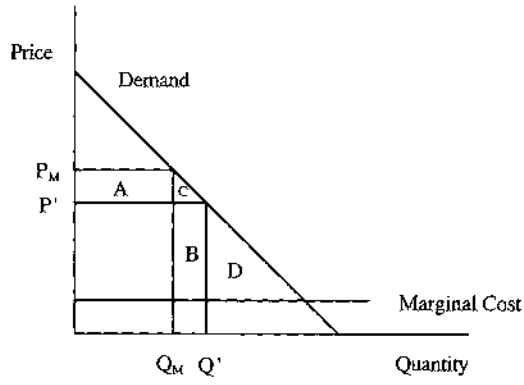
entee, this Article shows that a regime with some uncertainty and delay can produce this reward more efficiently than a regime in which enforcement is instantaneous and certain.

A. *The Intuition: Why It Is Possible to Restrict Patentees' Market Power Without Reducing Innovation Incentives*

Before demonstrating how uncertainty and delay restrict market power by inducing limited infringement, it is useful to explain why restricting patentees' market power is a worthy goal. Two intuitions suggest the worth of restricting patentee market power — one relating to fixed and the other to increased patent length. For convenience, we will refer to these intuitions as the "stationarity intuition" and the "Ramsey intuition."

The Stationarity Intuition. Small restrictions in the patentee's monopoly market power are efficient — even if these restrictions reduce the patentee's expected profits. The reason is that the last increment by which an unconstrained patentee chooses to increase price hurts society much more than it helps the patentee. Because an unconstrained patentee maximizes profits by choosing the quantity or price that reaches a flat point — what mathematicians call a stationary point — on the profit surface, small changes away from the profit-maximizing price or quantity will have only second-order effects on profits.[6] But small decreases in price will have first-order effects on the deadweight loss — so that reductions in price have larger impacts on welfare than profits.

FIGURE 1



6. The "envelope theorem" also exploits this aspect of the maxima (or minima) of functions. *See* George A. Akerlof & Janet L. Yellen, *Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?* 75 AM. ECON. REV. 708 (1985); Ian Ayres, *Pushing the Envelope: Antitrust Implications of the Envelope Theorem*, 17 MISS. C. L. REV. 21 (1996).

The intuition for this result can be seen in Figure 1. A reduction in the monopoly price from P_M to P' has two effects on the patentee's profits: the lower price reduces the profits on current sales (the reduction in profits is represented by the area of rectangle A), but the lower price also increases the quantity sold (from Q_M to Q') at a supra-competitive price (this increase in profits is represented by the area of rectangle B). So geometrically, the effect on profits of lowering the price below the monopoly level is:

$$\textit{Effect on Profits} = \textit{Area } B - \textit{Area } A$$

The patentee acting as a monopolist would choose a price where these two effects on profits were completely offsetting — but even reductions below the monopoly price cause a relatively small decrease in profits because the increase in profits from additional sales (Area B) almost offsets the decrease in profits from a lower price (Area A).

This reduction in profits is especially small in comparison with the increase in social welfare (consumer and producer surplus). As shown geometrically in Figure 1, the increase in welfare caused by lowering the monopoly price is:

$$\textit{Effect on Welfare} = \textit{Area } B + \textit{Area } C$$

From antitrust scholarship, it is well known that the patentee's loss of Area A profits does not reduce social welfare because what the patentee loses in this area, the consumer gains. But what is not often emphasized is that the effect on profits is much smaller than the effect on social welfare. Because of the offsetting profit rectangles, the effect on profits is *second-order* in comparison with the *first-order* effect on social welfare.

The next section will show in a simple linear example that reducing the monopoly price by 10% might only reduce the patentee's profits by 1%, but can reduce the social costs of monopoly by 19%.[7] Because the last bit of monopoly overcharging is so disproportionately damaging, restricting the patentee's monopoly power a small amount is likely to increase social welfare. The benefit of reducing the deadweight loss of supra-competitive pricing is likely to outweigh the costs of a slightly lower incentive to innovate.[8]

The Ramsey Intuition. While the stationarity intuition only pertains to small changes away from the profit-maximizing output, a

7. *See infra* Table 1.

8. *But cf. infra* text accompanying notes 36-37 (describing pathological example in which elasticity of innovation with respect to expected profit was extremely high).

second intuition applies to even larger reductions in market power. To wit: larger restrictions in a patentee's monopoly power are efficient, if the patent's length is increased to keep the patentee's expected profit constant.[9] If we hold patentees' profits constant by appropriately lengthening the patent life, consumers would be better off living under oligopolistic pricing for a longer period than monopoly pricing for a shorter period. The intuition for this trade-off between market power and duration can be seen as an implication of inter-temporal "Ramsey pricing."[10]

Ramsey pricing minimizes the distortionary effect (read: deadweight loss) of generating a given amount of tax revenue. The optimal tax structure for generating a given amount of revenue will tend to (1) tax as many goods as possible (so as to create small distortions in a broad variety of markets),[11] and (2) tax goods with inelastic demands more severely than goods with elastic demand. The first implication — regarding broad-based tax schemes — is particularly relevant to patent policy.

Allowing a patentee to charge a supra-competitive price is analogous to allowing a patentee to impose a tax — because both induce deadweight loss. The Ramsey intuition by itself suggests that government would want to place a small tax on all products to generate a given reward for a patentee. The problem with this approach is that government does not have sufficient information to know how much of a reward is appropriate for a particular product. Giving patentees the ability to charge monopoly profits allows the government to tie the size of the reward in a rough way to the size of consumer surplus created by the patent. Yet even if we limit

9. This will only be true if the patentee is risk-neutral and if preemptive innovation does not independently limit the patentee's ability to profit from the original innovation in later years. If patentees are risk-averse, patent duration will need to be lengthened more to maintain a constant incentive to innovate. The latter possibility of preemptive innovation is discussed below with regard to choosing the appropriate discount rate. *See infra* text accompanying notes 49-50.

10. *See* Frank P. Ramsey, *A Contribution to the Theory of Taxation*, 37 ECON. J. 47 (1927). For a review of the subsequent literature, see William J. Baumol, *Ramsey Pricing, in* 4 THE NEW PALGRAVE: A DICTIONARY OF ECONOMICS 49, 49-51 (John Eatwell et al. eds., 1987). *See also* Robert Cooter, *Optimal Tax Schedules and Rates: Mirrlees and Ramsey*, 68 AM. ECON. REV. 756 (1978); Stephen Law, *Inter-Temporal Tie-ins: A Case for Tying Intellectual Property* (Univ. of New Brunswick Working Paper 98-06, 1998); Frank Mathewson & Ralph Winter, *Tying as a Response to Demand Uncertainty*, 28 RAND J. ECON. 566 (1997) (arguing that a monopolist might also use tying to implement Ramsey pricing across products — instead of charging a monopoly price for the product it has a monopoly on and a competitive price for a related product).

11. This can be seen as an application of the stationarity intuition in another setting. The first increment of taxation has a second-order effect on social welfare but a first-order effect on tax revenue.

ourselves to reward structures that allow patentees to extract consumer surplus, it is better to spread the patentees' market power over multiple years.

If we view each year (or period) as a separate product, then the most efficient way — *à la* Ramsey — to produce a certain tax revenue (patentee profit) would be to impose some tax on every product. We wouldn't expect to tax only some of the goods. But an idealized patent regime violates the Ramsey intuition by taxing just the first twenty years and imposing no "tax" on subsequent years. If the periodic demand curve for the patented product remains constant as time passes[12] and becomes more elastic as price increases,[13] then it will be more efficient to raise the patentee's profits by taxing all periods equally than by imposing high taxes on some (early) periods and no taxes on other (later) periods.[14]

A central lesson of "marginalism" is that partial restraint of an externality may be much more efficient than either complete restraint or no restraint. But current patent law is curiously indifferent to this traditional economic concern. Under the idealized regime, the patentee would have unchecked monopoly power for the duration of the patent, and then no legally enforced market power for all succeeding years.

Our approach is consistent with the "ratio test" developed by Louis Kaplow.[15] Kaplow suggested that in "assessing the desirability of retaining a currently permitted practice" it is useful to com-

12. If the demand becomes more elastic over time, then raising most of the revenues in the first years may be efficient. But later years should still be "taxed." For example, if the demand curve is linear and the x-intercept decreases while the y-intercept remains constant, then Ramsey pricing would imply constant taxes over time. In other cases, if the demand curve's elasticity increases over time, policy makers will want lower but still positive amounts of "taxation" in the later, more elastic periods. In the probabilistic patents model, lower taxation in later years could be accomplished by awarding lower expected damages for infringement in later years. For the remainder of this Article, we restrict our attention to constant tax rates (i.e., constant expected damage rates) during the life of the patent.

13. Most standard demand curves — including all concave, linear, and not overly convex demand curves — satisfy this condition. *See* Klemperer, *supra* note 2, at 122. Linear demand curves become more elastic as price increases — even though the slope of the demand curve remains constant — because proportionate increases in price cause increasingly large proportionate reductions in demand. Monopolists will always increase price until they reach an elastic portion of the demand curve, and such a portion always exists for sufficiently high prices on demand curves that are linear.

14. The condition that demand becomes more elastic as price increases is stronger than necessary. It relates to the second implication of Ramsey taxation, that it is preferable to tax inelastically- rather than elastically-demanded goods. If demand is much more inelastic at higher prices, it might be preferable to tax a small number of periods at high prices rather than a larger number of periods at lower prices.

15. *See* Louis Kaplow, *The Patent-Antitrust Intersection: A Reappraisal*, 97 HARV. L. REV. 1813 (1984).

pare the patentee's incremental reward to the incremental social loss resulting from the practice in question, because "[i]n general, the higher the ratio, the more desirable the practice."[16] When applying Kaplow's test to patent pricing itself, one finds that allowing patentees to exploit full monopoly power produces an extremely low[17] and hence undesirable ratio — indicating that giving patentees the ability to raise price all the way to the monopoly level is not cost justified. The ratio test suggests that policymakers should analyze the incremental profits per dollar of social loss on various dimensions of patent scope. If one margin of protection produces lower patentee profits per dollar of social loss than another margin, it makes sense to reduce the patentee's entitlement where the ratio is low and to expand the patent entitlement where the ratio is high. Doing so can reduce the deadweight loss of patents without reducing the patentee's expected profits. The Ramsey intuition suggests that just this type of trade-off is possible with regard to monopoly pricing and duration.

B. *Implementation: How Probabilistic Patents Restrain Patentees' Market Power*

Probabilistic patents can constrain patentees' market power, because uncertainty and delay can induce limited amounts of interim "infringement."[18] If the probability that the patent will be enforced is sufficiently low, entrants may find it profitable to produce the patented product. The oligopolistic profit that the infringers expect to make when the patent is not enforced compensates for the fact that the infringers will have to more than disgorge these profits in order to make the patentee whole when the patent is enforced.

Infringement during the patent's life will tend to expand industry output and decrease the market price. This enhanced competition implements the goal of restraining the patentee's market power. The infringers' entry will be *limited*, however, because the marginal benefit of entry is weighed against pro-rata damages. As infringers help to drive down the market price, at some point additional entry will be deterred because the depressed market price

16. *Id.* at 1829 n.4, 1831. Kaplow's approach is applied to the copyright issue of "fair use" in William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659 (1988).

17. Indeed, at the stationary point — the profit-maximizing price — the ratio by definition is zero.

18. For convenience, the interim production by nonpatent holders is referred to as "infringement" — even though under a probabilistic regime, there is some probability that the "infringer" will not need to pay damages.

reduces the profitability of additional production and exacerbates expected damages. The amount of this limited infringement is determined by the certainty of enforcement. By changing the probability of enforcement, lawmakers can limit patentee market power by various degrees.

Initially, it might seem that using interim infringement to restrain industry pricing might substantially decrease the patentee's profitability — because the patentee must split the industry profits with the infringing entrants.[19] This intuition, however, ignores the fact that patentees expect damages from infringement. If potential infringers enter to the point where they expect to earn zero profits, then in expected terms the infringers will be paying all of their profits to the patentee.

The infringers in this stylized story are not pariahs, but perfect agents of society — restraining the excesses of monopoly pricing without personal profit. In many contexts, the infringing firms may expect to earn positive profits even after paying damages for infringement, but the social benefits of restraining the patentee's monopoly power may be sufficiently great that allowing profits for infringement may still enhance welfare. The next section develops a simple model to show how a probabilistic patent regime can reduce but not eliminate a patentee's monopoly power by inducing limited amounts of infringement.

I. A Simple Model

To formally capture the effects of uncertainty and delay, this section explores how market participants would behave under a stylized "probabilistic patent" regime. This regime has two unique features:

(1) a patent holder would only have a right to seek compensatory damages *after* the patent's expiration (patent holders could not seek injunctions prohibiting infringement before the patent's expiration); and
(2) a court would only award these damages with some probability less than one.

A probabilistic patent incorporates both imperfect enforcement and delay, because a patent holder cannot learn whether infringing producers will have to pay damages until the patent has expired, and this delayed decision is itself somewhat random. Thus, the

19. In terms of Figure 1, infringement might not only drive the price to P', but the patentee would be supplying only a portion of Q'.

model imagines a regime in which the patentee of a true innovation — that is, an innovation deserving immediate and certain enforcement under current law[20] — would have to wait until the day the patent expired to learn if a court would award damages for any past infringement; and the court would make this determination simply by flipping a weighted coin with, say, only an 80% chance of enforcement. If the coin lands on "enforce," the court would force infringers to "make the patentee whole" by increasing the patentee's profits to the monopoly level (what the patentee's profits would have been without infringement); if the coin lands on "don't enforce," infringers would not have to pay any damages.[21]

This model is intentionally reductive to illuminate how uncertainty and delay can produce higher welfare than an "idealized" patent regime, which provides immediate and certain resolution of validity disputes. While the thesis of this Article is not that judges should actually flip weighted coins in adjudication, understanding the perverse benefits of uncertainty and delay suggests that policymakers may not want to spend inordinate amounts of money to eliminate certain types of uncertainty and delay from the current system.

To show how uncertainty and delay in resolving patent disputes can perversely enhance welfare, consider the following algebraic model. Assume that a patent is awarded for a product that has a simple linear demand:

$$p = 100 - q \qquad (1)$$

where p = price; and

q = the total quantity sold per period.

Also assume constant marginal costs of production — which without loss of generality can be set equal to zero.[22] If the legal system affords the patentee the certain right to immediately enjoin

20. Thus our model only admits the possibility of Type I uncertainty (the possibility that valid patents will be unenforced) excluding the possibility of Type II uncertainty (the possibility that invalid patents might be enforced). We will later discuss the extent to which uncertainty in the real world gives rise to each of these two distinct possibilities. *See infra* Part III. For now, suffice it to say that we could imagine a regime that keeps the amount of Type II uncertainty constant but which increases Type I uncertainty. For example, one could imagine a regime in which a court or an agency immediately adjudicated the validity of the patent but then waited 20 years before flipping a coin to decide whether any damages would be paid for intervening infringement.

21. In either instance, the result of the coin flip should not affect future behavior, because the validity of the patent is only determined after the patent has expired.

22. If the product costs $10 per unit to produce, the demand schedule could simply be rescaled to produce the same results.

infringement,[23] then the patentee will maximize its profits by producing the monopoly output of fifty each period, and would earn a profit of $2,500 during each period that the patent was enforced. For simplicity, let's begin by assuming that the patent is valid for just one period.[24]

Now consider what would happen under a probabilistic patent regime in which the court did not decide whether to enforce a patent — in the limited sense of awarding damages for past infringement — until the end of the patent's life. Let w equal the crucial probability that a patentee will "win" a subsequent infringement suit to enforce the patent. Because the enforcement of the patent will only be resolved at the end of the patent's life, the legal rules must specify the ex post damages for interim (and potentially infringing) production by other entrants. The following assumption is made about these ex post damages:

> If the court decides (with probability $1 - w$) not to enforce the patent, the entrants producing before the court's decision owe no damages for infringement. But if the court decides (with probability w) to enforce the patent, then infringing producers must pay their pro-rata share of the damages that would be sufficient to make the patentee whole.[25]

For the linear demand assumed in equation (1), the total damages would need to be sufficient to raise patentee profits to the monopoly level of $2,500 that she would have earned without infringe-

23. Stiff damages or strong injunctive protection of a patent entitlement would constitute what Calabresi and Melamed have called a "property rule." *See* Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972). Recently, Robert Merges has written several pieces exploring the implications of property and liability rules in intellectual property. *See, e.g.*, Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293 (1996); Robert P. Merges, *Of Property Rules, Coase, and Intellectual Property*, 94 COLUM. L. REV. 2655 (1994) [hereinafter Merges, *Of Property Rules*].

24. It is straightforward to allow patents to last for multiple periods (and to discount future profits) — or to define a period as lasting for 20 years.

25. We further assume that infringing producers (possibly with the aid of litigation bonds) will be able to pay these damages, that is, they will not be able to evade their liability through bankruptcy, etc.

Instead of being liable pro-rata, one might imagine that infringers would only be held liable for the marginal impact of their infringement on the patentee's damages. Judge Frank Easterbrook has rejected this standard:

> [T]here is the question whether [the defendant] is entitled to "credit," as it were, for the price erosion caused by the other infringers' sales. . . . [T]he answer is no. . . . A judge would not let the infringers play a game of whipsaw, in which each argued that it should pay less of the damages than its share of the sales, because the price would have been depressed anyway even had it never infringed. That approach would lead to a less than compensatory award. Indeed it might lead to no damages at all, even though by hypothesis the patent holder and his licensee have been injured.

In re Mahurkar Double Lumen Hemodialysis Catheter Patent Litig., 831 F. Supp. 1354, 1392 (N.D. Ill. 1993).

ment. Thus, the damages of the *i*th entrant (conditional on the court ultimately deciding to enforce the patent) are assumed to be:

$$\left(\frac{q_e^i}{Q_e}\right)\ [\$2{,}500 - q_p\ (100 - q)] \qquad (2)$$

where q_e^i = the quantity produced by the *i*th entrant;
Q_e = the quantity produced by all entrants; and
q_P = the quantity produced by the incumbent patentee.

The model assumes that an arbitrarily large, competitive fringe of risk-neutral potential entrants is able and willing to enter and expand production if entry is profitable. It also assumes that the patentee is risk-neutral, and that if there is interim entry and production by nonpatented entrants, the patentee and the infringing entrants will strategically choose the quantity to produce — according to what in the literature is called a Cournot game.[26]

With these assumptions it is possible to derive the equilibrium outputs of both the entrants (Q_e) and the patentee (q_P) expressed solely as a function of the probability (w) that the court will enforce the patent:[27]

$$Q_e = \frac{(2 - 2w)\ 100}{(2 - w)} \qquad (3)$$

and

$$q_p = \frac{100w}{2\ (2 - w)} \qquad (4)$$

---

26. In Cournot models of oligopolistic competition, each competitor chooses a quantity to produce that maximizes its profits, given an equilibrium expectation about the output of other producers. w is assumed to be common knowledge. The order of play is: (1) the patent is awarded; (2) the patentee and competitive fringe simultaneously choose output levels; and (3) the court flips a coin and announces damages if the patent is to be "enforced."

The possibility of risk-averse patentees is discussed *supra* note 9.

27. Given the entrants' expectation about the patentee's quantity (q_P), the competitive fringe will enter and increase its output until the entrants' profits are equal to the entrants' expected damages from losing an infringement suit. Algebraically, this equality can be expressed as:

$$Q_e\ (100 - q) = w\ [\$2{,}500 - q_P\ (100 - q)]$$

Given the patentee's expectation about the quantity that the competitive fringe will produce (Q_e), the patentee will produce the quantity that maximizes its profits. Since its profits are $w[\$2{,}500] + (1 - w)\ [q_p\ (100 - q)]$ (and remembering that $q = q_P + Q_e$), a profit maximizing patentee will produce:

$$q_p = \frac{100 - Q_e}{2}$$

Solving these two equations yields the expressions for the patentee's and entrants' output shown in the text.

This example shows how probabilistic patents induce limited amounts of infringement. The output of entrants increases with uncertainty (that is, as the probability of enforcement, w, decreases). And the increase in the entrants' output more than offsets the decrease in the incumbent's output — so that total industry output $q = (Q_e + q_P)$ increases as well. But increased uncertainty does not induce unlimited entry because competition drives down the market price toward marginal cost, eventually reducing the benefits of supra-competitive pricing below the expected cost of paying pro-rata make-whole damages.

TABLE 1: EFFECT OF UNCERTAINTY ON PATENTEE'S PROFITS AND SOCIAL WELFARE

Probability of Enforcement w	Price	Expected Patentee Profits	Social Cost	% of Monopoly Profit	% of Monopoly Social Cost	Ratio of % S.C. to % Profit
100%	50.0	2500.0	1250.0	100.00%	100.00%	1.00
95%	45.2	2477.3	1023.2	99.09%	81.86%	0.83
90%	40.9	2417.4	836.8	96.69%	66.94%	0.69
85%	37.0	2329.9	682.9	93.19%	54.63%	0.59
80%	33.3	2222.2	555.6	88.89%	44.44%	0.50
75%	30.0	2100.0	450.0	84.00%	36.00%	0.43
70%	26.9	1967.5	362.4	78.70%	28.99%	0.37
65%	24.1	1827.8	289.8	73.11%	23.18%	0.32
60%	21.4	1683.7	229.6	67.35%	18.37%	0.27
55%	19.0	1536.9	179.8	61.47%	14.39%	0.23
50%	16.7	1388.9	138.9	55.56%	11.11%	0.20
45%	14.5	1240.9	105.4	49.64%	8.43%	0.17
40%	12.5	1093.8	78.1	43.75%	6.25%	0.14
35%	10.6	948.1	56.2	37.92%	4.50%	0.12
30%	8.8	804.5	38.9	32.18%	3.11%	0.10
25%	7.1	663.3	25.5	26.53%	2.04%	0.08
20%	5.6	524.7	15.4	20.99%	1.23%	0.06
15%	4.1	389.0	8.2	15.56%	0.66%	0.04
10%	2.6	256.2	3.5	10.25%	0.28%	0.03
5%	1.3	126.6	0.8	5.06%	0.07%	0.01
0%	0.0	0.0	0.0	0.00%	0.00%	

Table 1 shows the equilibrium price and the patentee's expected profits for various levels of enforcement uncertainty. For example, if w equals 75%, equations (3) and (4) predict that the patentee would produce 30 units (20 down from its monopoly output of 50) and that entrants would produce 40 units. The expanded industry production drives the market price down from the monopoly price of $50 to $30.[28]

28. In this model, the fact that the market price (in dollars) equals the patentee's output (in units) is an artifact of the particular linear demand function and not a general result.

As is well known, reducing the market price reduces the deadweight loss of monopoly. In this example, the reduction in social cost is significant: infringement eliminates almost two-thirds (64%) of the deadweight loss of monopoly pricing (down $800 from $1,250 to $450).[29] But the patentee's expected profits are reduced *by less than one-sixth* from the monopoly level.[30] In this example, the patentee expects to earn a total of $2,100: $900 on the units she produces ($30 per unit on 30 units) and $1,200 in damages (a 75% chance of $1,600). A crucial aspect of the competitive fringe assumption is that the profits from infringing production are on average all expected to go to the patentee.[31]

Uncertainty and delay are both necessary to induce limited amounts of interim infringement (and their salutary effects on the patentee's pricing). Delay is necessary, because a regime that immediately resolved any uncertainty about a patent's validity would recreate the all-or-nothing excess of the idealized regime.[32] Patentees would immediately learn whether or not they had complete monopoly power — so that with probability w the patentee would enjoin all interim infringement and set the full monopoly price. Uncertainty is necessary, because a regime that produced delayed, but certain, decisions (therefore refusing injunctions and only awarding damages ex post with probability w = 1), would also deter all interim infringement. Potential infringers would know that infringement would inevitably expand industry output above the monopoly level and thus reduce total industry profit. Accordingly, they would know that any profits they would earn from infringement in the interim would be insufficient to compensate the patentee for lost profits. Potential infringers, knowing that they could not break even, would not infringe.

---

29. This reduction in the deadweight-loss triangle is represented in Figure 1. At the monopoly price the deadweight loss equals the area of the triangle comprising Areas $B + C + D$, but the reduction in price reduces the deadweight loss triangle to Area D.

30. As shown in Table 1, profits fall $400 from $2,500 to $2,100.

31. The competitive fringe earns profits of $1,200 ($30 per unit on 40 units) but 75% of the time must pay damages of $1,600 — so that competition induces entry to the point that entrants expect to earn zero profits. On net, the entrants' 25% chance of earning $1,200 compensates for the 75% chance of losing $400.

32. Delay would not be necessary to induce limited infringement if (1) damages were not set to compensate the patentee for lost monopoly profits, and (2) prospective damages were possible instead of an injunction. Under such circumstances, the partially compensating damages would be equivalent to a compulsory license. *See infra* text accompanying notes 115-27.

Even though delayed judgement is needed to induce interim infringement, the law may not need to mandate delayed decision. As long as the patentee has the option of postponing judgement, the patent holder has a strong incentive to delay the court's decision. *See infra* text accompanying notes 66-67.

Table 1 shows that introducing small amounts of uncertainty reduces deadweight loss much more than it reduces the patentee's expected profit. By itself this observation is not a very powerful guide to policy because it does not tell us whether the benefit of lower prices is worth the cost of reduced innovation incentives. The next sections apply the stationarity and Ramsey pricing intuitions to the model to show why, with either fixed or enlarged patent durations, concerns about innovation need not undermine the usefulness of probabilistic patents.

A. *Introducing Uncertainty Without Changing the Patent's Duration*

The essence of the stationarity intuition is illustrated in the second-to-top row of Table 1. If the probability of enforcement w is 95%, the market price will be cut nearly 10% (from $50 to $45.20) and the social cost of the patent will fall more than 18%.[33] But near the point where the profit function is maximized — the point where profits are stationary — small changes in price have a negligible effect on profit. As shown in Table 1, this rather significant decrease in deadweight loss occasions less than a 1% drop in profits.[34]

While the stationarity intuition holds as a matter of theory only for arbitrarily small perturbations in price or quantity,[35] the intuition can hold as a practical matter for noninfinitesimal price reductions. Table 1, for example, shows that deadweight loss is much more sensitive than profits *even for noninfinitesimal reductions in price.* Moreover, this result holds for a wide variety of demand and cost curves. The example that we can eliminate 19% of the deadweight cost of patents while only reducing the patentee's expected profit by 1% suggests that policymakers should be willing to countenance small amounts of delay and uncertainty even without compensating extensions of the patent life. For it seems likely that society would be better off bearing the reduction in innovation caused by a 1% reduction in a patentee's expected profit, rather than bearing the additional deadweight loss caused by monopoly

33. Table 1 shows the social cost of $1,023.20 to be 81.86% of the monopoly social cost ($1,250).

34. Table 1 shows profits of $2,477.30 to be 99.09% of monopoly profits ($2,500).

35. As George Akerlof and Janet Yellen have noted with regard to the stationarity prerequisite for the envelope theorem: "[F]or the [envelope] theorem to have practical relevance, it must be true for finite values of ω, corresponding to economically noticeable shocks, and not just for infinitesimal ω . . . ." Akerlof & Yellen, *supra* note 6, at 711.

pricing.[36] Introducing discrete amounts of uncompensated uncertainty might reduce welfare if innovations were relatively sensitive to small reductions in price: It would be possible to construct pathological examples in which innovations were discontinuous in the patentee's profits so that even a 1% reduction in expected profit would be sufficient to deter innovation. But our patent law is not tailored to induce efficient innovation on a case-by-case basis — such tailoring would be prohibitively expensive. In our one-size-fits-all twenty-year regime, it is much more natural to model the responsiveness of innovation to changes in profit as a continuous function.

Thus, the stationarity intuition suggests that even if patent length is held constant, some small but discrete amounts of delay and uncertainty might be cost-justified. The Ramsey pricing intuition goes one step further, demonstrating that more substantial drops in patentee profits caused by uncertainty and delay — and the innovation reduction that those drops entail — can be offset by extending patent life. The next subsection shows how policymakers might extend patent life to hold the patentee's expected profits constant under a probabilistic regime.[37]

B. *Lengthening the Patent's Life to Maintain the Innovation Incentive*

For sufficiently large reductions in price, the stationarity intuition no longer holds true. Uncompensated decreases in expected profits might retard innovative effort enough to more than offset the benefits of reduced deadweight loss. But the range of cost-justified uncertainty can be extended by increasing the duration of the patent's life. Lengthening the duration of the patent's validity can enhance the patentee's expected profits and thereby offset the innovation effects of uncertainty.[38] This section calculates how much longer the patents would need to remain (probabilistically) valid to maintain a patentee's expected profitability. Lengthening the patent's validity to maintain the patentee's expected profits

36. *See* Tandon, *supra* note 2, at 474. The elasticity of social surplus with regard to R&D has been estimated to be approximately 0.10. *See id.* A mild reduction in profit would occasion a mild reduction in R&D and a correspondingly small reduction in value added from patents.

37. Obviously, such a lengthening analysis only has practical usefulness to the extent that statutory inertia can be overcome to actually increase the current patent duration. *See* Kaplow, *supra* note 15, at 1841.

38. To be more precise, in a probabilistic regime, lengthening the patent's duration is only lengthening the patent's *potential* validity, because after the patent expires a court may hold (with probability $1 - w$) that the patent was not valid.

eliminates the concern that probabilistic patents will retard initial innovation; patentees' incentives to innovate should be unaffected by a move toward a more uncertain, but longer-lived patent.[39]

By holding the patentee's innovation incentive constant, we can focus on whether longer-lived probabilistic patents enhance consumer surplus. It turns out that in the simple model (with constant annual demand) consumers are better off paying oligopolistic prices for an extended duration than paying monopoly prices for twenty years. To see why this is so in a world without discounting, look again at Table 1. If there is a 10% chance that the patent will not be enforced (w = 90%), then the resulting infringement will reduce profits per period to about 96.7% of what they would have been with certain enforcement. To maintain the patentee's expected profits, it would be necessary to lengthen the patent's duration to approximately 103.4% (1/.967) of its initial length.

But Table 1 also indicates that a 90% probabilistic regime reduces the *per period* deadweight loss of the patent to approximately 67% of what it would be in the idealized regime. Social welfare would be enhanced by increasing the patent length to compensate for the reduction in periodic profits because bearing 67% of the deadweight loss per period for a duration that is 3.4% longer reduces the total deadweight loss from monopoly to 69.2% (67% times 103.4%) of what it would be under an idealized regime. The incentives for innovation are unaffected because the patentee receives the same expected return, but the social inefficiency of supra-competitive pricing is reduced by more than 30%. The right-hand column of Table 1 — depicting the "Ratio of % Social Costs to % Profits" — thus shows what percentage of monopoly social costs would remain in a probabilistic patent regime that lengthened the patent's duration sufficiently to keep the expected profits constant.[40]

This analysis resonates with Louis Kaplow's proposed standard for judging whether compensating increases in patent duration improved social welfare: "[O]ne determines whether permitting the practice would impose more cost per unit of incremental reward than would result from lengthening the patent life to provide the

39. Recall that we have assumed that patentees are risk-neutral. *See supra* note 9.

40. The values for this column are calculated as follows: Generalizing from the example provided in the text, the optimal probabilistic regime requires multiplying the patent length by the inverse percentage of monopoly profit produced each period by the probabilistic regime. The overall percentage effect on social welfare will thus be the periodic effect on social cost ("% S.C.") multiplied by the percentage increase in the length of the patent (1/"% Profit"), which equals the right-hand column.

same reward."[41] In essence, policymakers should seek the lowest cost way of generating a particular level of profit. If raising the periodic price produces a higher social cost per dollars of profit than extending the life of the patent, then it is more efficient to extend the duration. Applying this standard to the example in which w equals 90%, we see that allowing the patentee to increase the market price from $41 to $50 (the monopoly level) produces approximately $5 of social cost for every dollar of profit generated for the patentee.[42] But, in sharp contrast, "lengthening the patent life to provide the same reward" produces only 35 cents of social cost for every dollar of patentee profit generated.[43] Thus, Kaplow's standard correctly captures the enhanced efficiency of a probabilistic patent regime in which the patent duration is lengthened to keep the patentee's expected profits constant.[44]

This simple example shows that consumers and the patentee might want to enter into a Coasean bargain to alter the idealized regime. In this bargain, the consumers would extend the patent duration in exchange for the patentee reducing its price below the monopoly level.[45] A reduction in price enhances periodic consumer welfare so much more than it reduces patentee profits, that increasing the patent length can compensate the patentee for reduced interim pricing and still leave extra surplus for consumers. This result holds not just for linear demand, but for any demand curve that is not overly convex — including all concave and constant elasticity curves. For such demand curves, the social cost per dollar of patentee profit generated decreases as the price is constrained further below the monopoly level.[46]

---

41. Kaplow, *supra* note 15, at 1829.

42. This calculation is based on $413.20 increased deadweight loss divided by $82.60 increased patentee profit.

43. For w = 90%, a compensating increase in patent duration increases the social cost by approximately $28 (= ($837/.9669) - $837). The social cost per dollar of profit is calculated by dividing the $28 increased deadweight loss by the $82.60 increased patentee profit.

44. For other examples of adjusting the patent length to keep the patentee's expected profits constant, see Gilbert & Shapiro, *supra* note 2; Klemperer, *supra* note 2, at 114; and Manfredi La Manna et al., *The Case for Permissive Patents*, 33 EUR. ECON. REV. 1427, 1435-37 (1989).

45. In essence, the consumers would be exchanging their entitlement to end the patent's validity (for a few years) for the patentee's entitlement to price at the monopoly level. This example suggests that a monopsonist consumer might have a strong incentive to enter into a contract with a longer duration than the patent is valid for in exchange for a lower price per period. The fact that buyers with market power engage in a particular activity can provide strong information for policy makers seeking to promote the welfare of more diffuse consumer groups. *See* Ian Ayres & John Braithwaite, *Partial-Industry Regulation: A Monopsony Standard for Consumer Protection*, 80 CAL. L. REV. 13 (1992).

46. *See* Klemperer, *supra* note 2, at 120-24. This is always true at sufficiently competitive prices, because (by stationarity) small increases in price affect patentee profits much more

This suggests that the probabilistic patent result might apply not only to many demand curves but also to many increments along a given demand curve. The intuition behind broad-based Ramsey taxing is that consumers will want to trade reduced pricing for a compensating increase in length under a broad array of demand conditions. Thus, the efficiency of restricting patentees' pricing applies not only to "local" reductions away from the monopoly price, but also to more substantial restrictions in the patentee's market power. Indeed, inspecting Table 1, one finds that continually lowering the probability of enforcement increases social welfare — even after the patent length is increased to keep the patentee's expected profits constant. This can be seen by inspecting the falling percentages in the far right-hand column. As described above, the ratio in this column shows the percentage of monopoly social cost that would remain in a probabilistic patent regime in which the patent's duration was increased to keep patentees' expected profits constant.[47] For example, if courts only enforced patents half of the time (w = 50%), there would be substantial amounts of infringement and the patentee's expected profit per period would only be 5/9 (55.56%) of what it would be if the patentee could charge the monopoly price. To keep the patentee's expected profits constant, the patent duration would need to be lengthened by 80% (that is, multiplied by 9/5). But Table 1 also shows that the deadweight loss per period in this equilibrium is just 1/9 (11.11%) of what it would be if the patentee charged the monopoly price. Once again, the ratio in the right-hand column shows the combined effect of an increased patent length and lower prices on the total deadweight loss: bearing just 1/9 of the periodic inefficiency for 9/5 periods produces a total inefficiency that is only 1/5 (that is, 20% as shown in Table 1) of the deadweight loss of the idealized monopoly regime.

---

than they affect social welfare. But it is also true at all prices if the elasticity of demand is nondecreasing, since a firm selling at a higher elasticity earns lower profits relative to the consumer surplus dissipated.

47. As a formal matter, policymakers would want to continue to reduce price (and lengthen the patent's duration) as long as:

$$\frac{\partial\left[\frac{SC}{\pi}\right]}{\partial p} > 0,$$

where SC and π are the expected periodic social cost and patentee profit, respectively. For the linear model set out above, this relevant derivative is always positive, indicating that compensated reductions in price all the way down to the competitive level are cost justified. This derivative will be positive for all levels of w for all demand curves that are not too convex.

The dominance of increasingly uncertain enforcement at all points along a demand curve that is not too convex heightens the perversity of the result, because without more it suggests that society would be best served if patents lasted an arbitrarily long time, but the law granted damages for infringement after the fact with an arbitrarily small probability. This result by itself proves too much, because it does not contain a limiting principle.

There are, however, two factors that limit the ability of lawmakers to lengthen the patent duration to compensate the patent holder for lower profits per period. First, the discounting of future profits limits the ability of lawmakers to sufficiently lengthen the duration in order to hold the present value of the patent holder's profits constant. A probabilistic regime still reduces the present value of the social inefficiency, but discounting forces lawmakers to add an increasing number of years to the patent's duration to keep the present value of the patentee's profits constant.

TABLE 2: COMPENSATING INCREASE IN PATENT DURATION AND NET EFFECT ON EFFICIENCY (ASSUMING THE REAL DISCOUNT RATE IS 7% AND STATUS QUO DURATION OF 20 YEARS)

Probability w	Number of Years	Price	Present Value of Patentee Profits	Present Value of Social Cost	Percentage of Status Quo Social Cost
100.00%	20.0	$50.00	$26,907.25	$13,453.63	100.0%
95.00%	20.4	$45.24	$26,907.25	$11,113.86	82.6%
90.00%	21.6	$40.91	$26,907.25	$ 9,314.05	69.2%
85.00%	23.6	$36.96	$26,907.25	$ 7,886.61	58.6%
80.00%	26.9	$33.33	$26,907.25	$ 6,726.81	50.0%
75.00%	32.5	$30.00	$26,907.25	$ 5,765.84	42.9%
70.00%	45.1	$26.92	$26,907.25	$ 4,956.60	36.8%
66.97%	∞	$25.17	$26,907.25	$ 4,525.50	33.60%

Table 2 shows the number of years by which it would be necessary to increase an idealized 20-year monopoly regime if the annual real discount rate is 7%.[48] Without discounting, Table 1 suggests that to

48. The total patent duration, L, required can be derived by setting the present value of receiving monopoly profits (π^M) for 20 years equal to the present value of receiving constricted profits ($\pi(w)$) for L years:

$$\pi^M \frac{(1-(1+\delta)^{-20})}{\delta\,(1+\delta)^{-1}} = \pi(w)\,\frac{(1-(1+\delta)^{-L})}{\delta\,(1+\delta)^{-1}}$$

which can be solved for L in terms of w and the other structural parameters in the model:

$$L = \frac{-\log\left[1-\frac{\pi^M}{\pi(w)}\,(1-(1+\delta)^{-20})\right]}{\log(1+\delta)}$$

The present value of social cost is simply the social cost from Table 1 discounted for L years.

compensate for a probabilistic regime in which w equals 70%, patent length would need to be increased approximately 27% (= (1/.787) - 1). Using a 20-year benchmark, this would mean an increase to approximately 25.4 years. But with 7% discounting, it would be necessary to increase the patent's length to more than 45 years. Indeed, because a perpetuity has finite present value, even making the patent's validity perpetual would not compensate for sufficiently large decreases in the patent holder's profits per period. In the current example (with the model's linear demand and a 7% discount rate), compensating increases in patent duration are not possible if the probability of enforcement, w, is set below 67%.

Second, the threat of preemptive innovation may also limit the ability of lawmakers to compensate lower periodic profits with longer patent validity. If technological advances are likely to make any particular invention outmoded after a small number of years, then extending the patentees' rights may not increase their expected return. Preemptive innovation would radically reduce the demand for the product in future periods. Because the discount rate can also be interpreted as the probability that the patent will become obsolete,[49] it is possible to account for this technological limit by increasing the discount rate. For example, if the risk of technological obsolescence made a 10% discount rate more appropriate, then it would be impossible to support enforcement levels of less than 73%.[50]

Another way to approach the practical impact of extending patent duration is to consider how much of a price reduction a given increase in patent duration could support. As a heuristic exercise, we consider an extremely stylized analysis of piroxicam, a popular antiarthritic drug that, until the patent expired in April 1992, was manufactured exclusively by Pfizer under the patent brandname "Feldene." In the year prior to the patent's expiration, Pfizer sold approximately 1.4 million units (100 20-mg. tablets) at $219 per

49. *See, e.g.*, Ian Ayres, *How Cartels Punish: A Structural Theory of Self-Enforcing Collusion*, 87 COLUM. L. REV. 295, 315 (1987).

50. This lower limit on enforcement levels is calculated by determining what fall in periodic profitability could not be compensated for by making the patent duration perpetual. Of course, a "perpetual probabilistic" patent as a limiting case would be impossible to implement because the court's decision whether to award damages for "interim" infringement would only come at the end of time. But a similar perpetual effect might be implemented with our "partial damages" regime, discussed *infra* section IV.B. Perpetual patents would violate the Constitution which gives Congress the power to reward exclusive rights for only a "limited" duration. *See* U.S. CONST. art. I, § 9, cl. 8.

unit.[51] But after the patent expired, generic manufacturers soon cut the price to $144 per unit.[52] If we assume that (1) the generic price equals the constant marginal cost; (2) the patented price represents the monopoly price; (3) the periodic demand for Feldene is linear and constant across periods; and (4) patentee profit equals zero after the patent expires,[53] then it is straightforward to assess the effect of trading longer duration for lower prices so as to keep the patentee's profits constant.

TABLE 3: HEURISTIC EFFECTS ON SOCIAL COST OF LENGTHENING FELDENE PATENT DURATION (BEYOND HYPOTHETICAL 20-YEAR PERIOD) (USING A 7 PERCENT DISCOUNT RATE)

Additional Years	Reduced Price Which Keeps Patentee's Profit Constant	Gross Reduction in Present Value Monopoly Social Cost ($ millions)	Percentage Reduction in Present Value of Monopoly Social Cost
1	207.8	146.0	26.0%
3	200.6	221.4	39.4%
5	196.4	260.4	46.3%

Table 3 shows that even modest increases in patent length can allow sufficient price reductions to substantially reduce the deadweight loss of monopoly. While this stylized example abstracts away from many important aspects of reality, the exercise underscores in a real world setting that nontrivial improvements in efficiency may flow from a regime that trades longer duration for reduced market power.

C. *The Government Has Sufficient Information to Implement Probabilistic Patents*

All problems of patent regulation ultimately devolve to a question of government information. If government had good information about the cost and/or consumer value of a particular

51. *See* James Kim, *Drug-company backlash*, USA TODAY, Nov. 4, 1992, at B1; conversation with Pfizer representative Roberta Lombardini (Nov. 3, 1995).

52. Conversation with Pfizer representative, Roberta Lombardini (Nov. 3, 1995).

53. Each of these assumptions is literally false. The periodic demand curves for patented products do not remain constant. And usually after a pharmaceutical patent expires, the patentee — when faced with generic competition — raises its price in order to focus on the relatively price insensitive segment of demand. *See, e.g.*, Ernst R. Berndt et al., *Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs*, *in* BROOKINGS PAPERS ON ECONOMIC ACTIVITY: MICROECONOMICS 133, 151-52 (Martin Neil Baily et al. eds., 1996); *cf.* Zvi Griliches & Iain Cockburn, *Generics and New Goods in Pharmaceutical Price Indexes*, 84 AM. ECON. REV. 1213, 1215 (1994) (stating that "the incumbent usually does not respond to entry by reducing its price").

innovation, it could pay efficient lump-sum bounties to the inventor and be done with the ex post deadweight loss altogether. Indeed, the Ramsey broad-based tax intuition itself suggests that the ex post distortion would be minimized if the patentees were able to impose a small tax on every conceivable product sold. Such a scheme is analytically equivalent to a lump sum reward funded out of the general fisc. The problem, of course, is that the government does not have sufficiently strong information to determine how to appropriately reward innovation.

The current patent system — which grants the patentee a monopoly for a limited number of years — is structured to minimize the government's information requirement. Specifically, the current system economizes on the patentees' private information. The government doesn't need to estimate the cost or value of innovation. Instead, the government allows the patentee to capture a rough and limited proxy of consumer value and then lets the potential innovator decide whether the benefits of innovation justify the costs.[54]

The benefits of probabilistic patents could also be created by a compulsory license regime that directly restricted the price that the patentee could charge (possibly in exchange for a longer duration). But compulsory license regimes face the same government informational hurdles as regimes with lump-sum bounties. It is particularly difficult for the government to estimate the demand and cost curves ex ante and to choose a price that would represent a certain percentage reduction in the monopoly markup.[55] And if the government did know enough to set an efficient compulsory license ex post the invention but ex ante the production, then it could go further and implement a first-best lump-sum reward.

Our argument is that lawmakers can improve on the current regime without knowing the exact shape of the demand and cost

54. There is evidence, however, that the returns to the patentee under the current system may capture much less than the social value of the patent. *See, e.g.*, Timothy F. Bresnahan, *Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services*, 76 AM. ECON. REV. 742, 753 (1986); Michael Kremer, *Patent Buyouts: A Mechanism for Encouraging Innovation*, 113 Q.J. ECON. 1137, 1141 (1998); Edwin Mansfield et al., *Social and Private Rates of Return from Industrial Innovations*, 91 Q.J. ECON. 221, 234 (1977).

55. Compulsory license regimes have been notoriously vague about the standards used to set "reasonable" royalties. *See* F.M. SCHERER, THE ECONOMIC EFFECTS OF COMPULSORY PATENT LICENSING 44 (Center for the Study of Fin. Indus., N.Y. Univ. Monograph Series in Fin. and Econs. No. 1977-2, 1997). Most often, however, the license fees are set to yield a reasonable profit above cost — instead of offering some percentage of what monopoly profit would be. Section IV.B will show that the benefits of probabilistic patents could also be obtained from a regime that provided patentees with certain enforcement but only awarded a percentage of the current make-whole damages.

curves. We can retain a system that ties the patentees' reward to the amount that can be extracted from consumers without giving patentees absolute monopoly power. To implement a probabilistic modification to the current regime, the government only needs to know (1) how to trade-off patent certainty for duration; and (2) the maximum uncertainty that can feasibly be offset with additional patent life. Theory suggests a simple answer to the first question, and empiricism is available to answer the second.

Trading Off Certainty and Length. Even though lawmakers would need to know the shape of the demand and cost curves to calculate the increase in duration needed to compensate exactly for an increase in uncertainty (and the price reduction that it creates),[56] a workable alternative is for the government to increase the duration in inverse proportion to the decrease in certainty. Multiplying the duration by the reciprocal of the probability of enforcement ($1/w$) typically offers a good approximation of the exact increase needed to assure that the patentee's expected profits are at least as great as under an uncompensated monopoly regime. Thus, under this approximation, if the probability of enforcement were 50%, the patent duration would be increased to 200% (1/.5) of its current level. Indeed, this simple reciprocal formula provides not only a workable approximation of the exact compensating increase in duration, but also guarantees patentees a slight increase in expected reward.[57] In our foregoing model with w equal to 50%, this approximation would actually increase the patentee's expected profits by 11% (from $2,500 to $2,777.80) but would decrease the social cost of the patent 77.8% (from $1,250 to $277.80).[58]

The Minimum Certainty that is Potentially Offsettable. While making the duration increase inversely with uncertainty neatly solves the government's trade-off problem, policymakers would still

56. The total patent duration necessary is the standard patent duration (20 years) times the reciprocal of the periodic percentage of monopoly profit, *see, e.g., supra* Table 1, which will generally turn on the particular shape of the demand and cost curves. *See supra* note 40.

57. To see this, let L equal the proportionate increase in patent length needed to keep the patentee's expected profits at the monopoly level. With this compensating increase in duration, the patentee will expect to earn:

$$\pi_p = L[q_p p + w(\pi^M - q_p p)]$$

where π_p = the patentee's expected profit and π^M = the monopoly profit. If the government sets $L = 1/w$, this expression for the patentee's expected profits can be simplified to show that it is slightly greater than monopoly profits:

$$\pi_p = \pi^M + \left(\frac{1}{w} - 1\right) q_p p > \pi^M.$$

58. Note, however, that demand and cost curves can be constructed such that this simple formula overcompensates the patentee to the extent that total social cost is increased.

need a limiting principle to decide how much of a periodic price reduction could be compensated for with an extended patent life. In a world with discounting and an increasing probability of technological obsolescence, lawmakers would need to estimate how many years the patent length would need to be extended to compensate for, say, a 90% probability of enforcement. But this calculation can be largely independent of estimates about the demand and supply conditions — thus easing the information requirement considerably.

To workably calculate the maximum compensable reduction in certainty, the government needs chiefly to know the rate at which the value of a patent generally decays.[59] And while the government assuredly does not have this knowledge for individual patents, estimates of just such decay rates are available for distinct patent classes. For example, Mark Schankerman has recently estimated the patent value decay rate using evidence from decisions on whether to pay yearly European renewal fees.[60] The data suggest that there are sharp distinctions in the rate at which the value of different types of patents decay. The value of pharmaceutical and chemical patents tend to decay at a 3% or 4% rate, while the value of mechanical and electronic patents tend to decay more quickly, at rates ranging between 10% and 15%. These data suggest that probabilistic patents are more likely to be useful for pharmaceutical and chemical innovations where there is a higher probability that the patent will still be valuable at the end of twenty years.[61]

---

59. In a simple model, imagine that the government knew that the probability that a patent would become obsolete in a particular year was δ (which as discussed above can be interpreted as taking into account both the time value of money and the probability of obsolescence). To compensate for the reduction in certainty, the government would want to set the patent length so that the present value of the probabilistic patent was equal to the present value of the nonprobabilistic 20-year patent. Instead of setting the patent length equal to the simple reciprocal of w, in a world with discounting it is necessary to set the approximate patent length equal to:

$$L_{approx} = \frac{-\log\left[1 - \frac{1}{w}\left(1 - (1+\delta)^{-20}\right)\right]}{\log(1+\delta)}$$

See supra note 48. As in *supra* note 57, it can be shown that this approximation produces slightly higher expected payoffs than does the monopoly regime.

60. *See* Mark Schankerman, *How Valuable is Patent Protection? Estimates by Technology Field*, 29 RAND J. ECON. 77 (1998); *see also* Ariel Pakes, *Patents as Options: Some Estimates of the Value of Holding European Patent Stocks*, 54 ECONOMETRICA 755 (1986); Ariel Pakes & Mark Schankerman, *The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources*, *in* R&D, PATENTS AND PRODUCTIVITY 73 (Zvi Griliches ed., 1984).

61. Schankerman, *supra* note 60, at 84.

The minimum sustainable level of certainty is a function of the decay rate. For example, in a simple model where demand for a patent is constant while the patent is nonobsolete, but where there is a δ probability that the patent will become obsolete, it is possible to show that the minimum compensable level of certainty (w^{min}) is:

$$w^{min} = 1 - (1+\delta)^{-20} \qquad (5)$$

This minimum w represents the degree of uncertainty that could be compensated by an infinite patent duration. If δ equals 10%, the minimum compensable w would equal approximately 85.1%. The government could use its knowledge about δ in this or a similar equation to estimate the lowest compensable amount of certainty. Alternatively, if as a pragmatic or political matter the policymaker were limited to a patent duration of L^{max} years, then the minimum compensable level of certainty (w^{min}) that would guarantee the patentee no reduction in the present value of expected profit would be:

$$w^{min} = \frac{1 - (1 + \delta)^{-20}}{1 - (1 + \delta)^{-Lmax}} \qquad (6)^{62}$$

The amount of uncertainty is simply a function of the patent decay rate (a variable that has been at least crudely estimated for different classes of patents) and the maximum practicable patent duration. A more complicated model could undoubtedly generate more complicated expressions turning on more variables, but the important point is that a government policymaker with limited information might still be able to have some idea of how much uncertainty to permit in exchange for a particular increase in patent life.

Indeed, the lawmaker might go further in harnessing the private information by giving the patentee the option of trading longer patent duration for less certainty. Under this regime of patentee choice, the patentee could opt for the current twenty-year protection with monopoly power or choose to extend the duration in return for a lower probability of enforcement. In essence the patentee would be given the choice of a schedule based on equation (6) and would be asked to pick a patent length and corresponding probability of enforcement. Patentees who believed their patent were unlikely to have value after 20 years would be unlikely to opt for longer durations, but an innovator with a patent whose value was particularly long lived might prefer to trade a longer patent duration for a lower probability of enforcement and the limited infringement that it is likely to engender. Even if only the innovators

62. This equation can be derived by re-solving the equation in note 59 in terms of w.

with the most valuable patents choose longer durations, substantial efficiency gains might be made, because a small number of the most valuable patents are responsible for a disproportionate percentage of total patent value.[63] Of course, patentees would choose strategically, and therefore the patentees' choice would not coincide with the social optimum. Indeed, in some examples the patentee would choose a probability of enforcement that even reduced social welfare.[64] For this reason, we ultimately favor the government choosing the level of uncertainty based on the government estimate of the rates of decay, as discussed above. But for those who give paramount importance to maintaining innovative incentives, a probabilistic regime with patentee choice would assure that patentees' expected rewards would not be decreased, because patentees could always opt for the status quo protection. While we do not seriously propose giving patentees the option of trading off a lower certainty of enforcement for a longer patent duration, the principle could be applied to more practical policy issues.[65]

D. *Uncertainty Creates Delay*

Our plan requires delay and uncertainty, but uncertainty by itself can give both the patentee and the infringer incentives to delay a court's determination. The patent holder has a strong incentive to delay the court's decision because if (with probability w) the patent is enforced, the patentee earns identical profits whether or not there is delay — but if (with probability $1 - w$) the patent is not enforced, the patentee earns much less profit from an immediate decision. If the court decides immediately that the patent will not be enforced, noninfringing entry could eliminate all of the patent holder's expected profits; but if the court delays its decision that the patent is not enforceable, then the patent holder earns oligopolistic profits from competing against more limited "infringement."[66]

63. Schankerman estimates: "The top 1% of patents account for 12% and 14% of the total value of patent rights in pharmaceuticals and chemicals, respectively, and 21% and 24% for mechanical and electronics patents (excluding Japan)." Schankerman, *supra* note 60, at 94.

64. For example, if a patentee believes that a patent will not become marketable until more than 20 years in the future, she might pick a low w (say 50%) that creates more deadweight loss than society would experience under the current regime.

65. For example, with regard to the Hatch-Waxman amendments, pharmaceutical patentees might be allowed to trade off extended patent duration for weaker preliminary injunction rights. *See infra* text accompanying notes 93-96.

66. However, if patent damages are not expected to make the patentee whole, the patentee may prefer to have an immediate decision. *See infra* note 116.

The model also suggests that the nonpatent holders may not have a strong incentive to oppose delaying tactics of the patentee. If (as the simple model assumes) nonpatent holders form a competitive fringe of potential supply, they should, before entering, be indifferent between immediate and delayed decision because they expect to earn zero profits in either case. If there are fixed costs of entry (or some other factor that might allow potential entrants to earn positive profits if the patent were held not valid), then these potential entrants might seek a declaratory judgement that the patent is not enforceable or might prefer delay, depending on the precise size of the fixed cost: immediate decision reduces the expected damages from infringement (because infringing production can be stopped), but also reduces the expected profits from noninfringement (because additional competitors will enter the market). If delay necessarily increases the litigation costs, either the patentee or the nonpatent holder may prefer immediate decision; but absent significant litigation costs *attributable to delay*, it is possible that delay will endogenously arise in a regime with uncertain enforcement — as patentees may have strong incentive to delay and the nonpatentees may not have a strong incentive to oppose such delay. Since many patent regimes already create substantial uncertainty about whether valid patents will be enforced, our model may help explain why infringement litigation takes so long to conclude.[67]

II. Caveats

This section explores a number of the most important reasons why the benefits of uncertainty and delay extolled in the foregoing model may not be achieved or might be outweighed by additional costs. "Probabilistic" patents may induce larger costs if they exacerbate the expense of litigation, interfere with follow-on innovation, or if infringing firms pay excessive fixed costs. And the projected benefits attributed to uncertainty — in restraining patent holders from raising prices all the way to the monopoly level — may be overstated if independent forces restrain patent holders from setting monopoly prices, or if price discrimination independently reduces the distortionary effect of monopoly power. While this is not a complete list of caveats,[68] it gives a flavor of some of the most

67. A particularly notorious example of protracted litigation concerns an interference proceeding that was declared in 1958 and resolved by the District Court only in 1980. *See* Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 Colum. L. Rev. 839, 901-02 & n.292 (1990).

68. For example, if the patentee is risk-averse, it would be necessary to have a longer compensating extension in duration to compensate for the risk that the patent would not be

important countervailing factors when considering uncertainty and delay in patent enforcement.

A. *Exacerbating Litigation Costs*

The simple model outlined above unrealistically excludes any costs of litigation. In the real world, infringement litigation carries with it significant costs.[69] It is commonly argued that uncertainty increases the cost of litigation. But this is not necessarily true. Rules that limit the resources spent on discovery or adjudication might increase uncertainty (because the decisionmaker would have poorer information) and simultaneously reduce litigation costs. For example, reforms that limit the amount of discovery and/or limit an infringement trial to two days at the end of the patent's life would probably lead to more uncertainty and lower litigation expense.[70] Some forms of uncertainty and delay would undoubtedly lead to litigation costs that swamp the benefits of limited interim infringement, but the point here is that legal policy could move toward procedural rules that increased or retained elements of uncertainty without enriching our siblings at the bar.

B. *Probabilistic Patents Might Hinder Follow-On Innovation*

The previous analysis might also understate the dynamic costs related to follow-on innovation.[71] If a patent's life is lengthened under a probabilistic regime to keep the patent holder's expected profits constant, then this lengthening might further block the emergence of valuable follow-on invention. While this criticism may be valid, it must be weighed against the fact that a probabilistic regime will also facilitate limited amounts of interim infringement

---

enforced. And while our model assumed that competition among potential infringers caused all the profits from infringement to on average be paid to the patentee by way of damages for infringement, one can construct models in which infringers earn profits in equilibrium (even after accounting for expected damages). Having infringers syphon off part of the industry's profits would also require larger compensating increases in duration. If either the risk-aversion or the infringer-profit effects were sufficiently pronounced, a probabilistic regime would not enhance welfare.

69. *See* F.M. SCHERER, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 454 (2d ed. 1980) (estimating total annual patent system administration and litigation costs to be roughly $300 million in 1978); Josh Lerner, *Patenting in the Shadow of Competitors*, 38 J.L. & ECON. 463, 470 (1995) ("[T]he patent litigation within USPTO and the federal courts begun in 1991 will lead to total legal expenditures (in 1991 dollars) of about $1 billion, a substantial amount relative to the $3.7 billion spent by U.S. firms on basic research in 1991."); Leslie Scism, *Insurance Helps Little Guy Sue Patent Infringer*, WALL ST. J., Nov. 25, 1996, at B1 (observing that patent litigation can "cost hundreds of thousands of dollars").

70. We will return to the question of whether the law can induce the right type of uncertainty, *infra* Part III.

71. *See* Scotchmer, *supra* note 1.

— including potential infringement from follow-on inventors.[72] It is accordingly possible that probabilistic regimes — like compulsory licensing regimes — might reduce the dynamic inefficiency of the idealized injunctive regime.[73]

C. *Infringers' High Costs May Dissipate Probabilistic Benefits*

Our model assumed that the infringers' entry into the market was efficient. Competition among potential infringers caused them to be perfect agents for society. They mitigated the deadweight loss by competing down the monopoly price, and turned over their profits (on average) to the patentee. If, however, there are fixed costs of entry or exit (or if entrants have higher marginal costs than the patentee), the infringers will also turn some of the patentee's profits into extra costs, which are a new kind of deadweight loss. In these cases, excessive entry may therefore increase social costs.[74]

Nevertheless, when entrants' additional costs make free entry a social bad, a correctly chosen w (and a compensating increase in patent duration) can both reduce social cost during the original patent life and reduce it by excluding unwanted excess entry during the extended patent duration. For example, imagine in our foregoing model (without discounting) that the fixed costs per period are 100. Free entry implies that nine firms (including the patentee) will enter, creating combined deadweight loss and fixed costs of 950. With these fixed costs, however, it is socially optimal for only four firms to compete.[75] Setting w equal to 3/7 and doubling the patent length to make the patentee whole produces the efficient amount of entry (three firms enter to challenge the incumbent patentee) and reduces social cost in both the initial patent period and the exten-

72. *See infra* text accompanying note 106 (discussing how broadening the doctrine of reverse equivalence may facilitate follow-on innovation).

73. Indeed, Professors Green and Scotchmer have shown that creating uncertainty about follow-on innovation may increase the expected returns of the initial patentee. *See* Green & Scotchmer, *supra* note 1. When it is certain that follow-on innovation will infringe, the follow-on innovators may credibly threaten not to invest in innovation (foreseeing that they will have to pay a large licensing fee). But Green and Scotchmer show that uncertainty about whether a subsequent innovation will be infringing may mitigate this threat of noninnovation and may thereby actually increase the original patentee's expected profits. *See id.* at 27; *see also* Scotchmer, *supra* note 1; Chang, *supra* note 1.

74. This is an example of the general proposition that with fixed costs (or if entrants have higher marginal costs), entry restrictions may be socially beneficial. *See* Jeremy I. Bulow et al., *Multimarket Oligopoly: Strategic Substitutes and Complements*, 93 J. POL. ECON. 488, 504-05 (1985); N. Gregory Mankiw & Michael D. Whinston, *Free Entry and Social Inefficiency*, 17 RAND J. ECON. 48 (1986).

75. That is, the socially optimal number of firms is four, assuming the foregoing Cournot model as set out in equations (1) and (2). Obviously, an all-powerful and all-knowing regulator could do better than permit Cournot competition.

sion period. Indeed, in this example, granting a perpetual patent (with $w = 3/7$) would increase the patentee's expected profits and reduce social cost by forever constraining the deadweight loss of excessive entry. Surprisingly, then, fixed costs do not generally militate against the use of probabilistic regimes. Fixed costs generally imply that the optimal amount of entry lies between monopoly and free entry levels. Probabilistic patents can be a means of inducing this intermediate level. Fixed costs reduce the benefits of constraining the monopoly price, but they simultaneously reduce the costs of extending the patent duration (because the extension constrains excessive entry in what would otherwise be an unpatented period). Since fixed costs reduce both the costs and the benefits of introducing uncertainty with compensating extensions in patent duration, the presence of fixed costs do not clearly contraindicate the use of a probabilistic patent regime. (Of course, we must be more cautious about encouraging infringement if we cannot simultaneously extend patent lives).

A similar analysis applies if infringers have higher marginal costs of production than the patentee: the benefits of infringement during the original patent period are reduced (or conceivably even reversed) but the costs of extending the patent life are likewise reduced (or possibly reversed).

D. *Existing Deadweight Losses May Not Be Severe*

If the patentee is able to perfectly price discriminate, then probabilistic patents would not be able to reduce the deadweight loss of monopoly pricing (because perfect price discrimination does not create any deadweight loss to begin with). However, although patentees as a legal matter are allowed to price discriminate, imperfect information about consumer valuations and an inability to stop resale often prevents patentees from engaging in any more than very crude forms of price discrimination. So probabilistic patents might still usefully be used to increase output.

The benefits of probabilistic patents may also be overstated if structural factors independently induce patent holders to charge less than the static monopoly price. For example, industrial organization theory suggests some such factors that might lead manufacturers with complete monopoly power to charge less than what otherwise would be the static profit-maximizing price, including:

producer learning curves, network externalities, and consumer switching costs.[76]

However, factors such as these do not affect the stationarity intuition that a small reduction from full monopoly power has only a small effect on a patentee's profits, but has a large impact on social welfare. The patentee will still choose a price (or quantity) that is profit maximizing, and while the price that maximizes profits in a more dynamic model may be below the profit-maximizing price in a simple, static model, it will still represent a stationary point on the patentee's profit function. These other factors also do not affect the Ramsey intuition that a regime of certain patents can be improved upon by probabilistic but longer-lived patents, though they may affect the quantitative significance of our results.[77]

Things are more complex if the patentee is engaged in oligopolistic competition with the producers of noninfringing differentiated products. In this case, facilitating limited infringement may not only affect the overall output of the patented product, but may also affect the output of the noninfringing products. While profits stolen by the infringers are, in equilibrium, returned to the patentee as expected damages, any profits transferred from the patentee to the noninfringing producers will not be recouped by the patentee.[78] Under such circumstances, the stationarity intuition would accordingly no longer hold.[79] *If* limited infringement induces a significant

---

76. *See, e.g.*, SCHERER, *supra* note 69; Alan Beggs & Paul Klemperer, *Multi-period Competition with Switching Costs*, 60 ECONOMETRICA 651 (1992); Darius W. Gaskins, Jr., *Dynamic Limit Pricing: Optimal Pricing under Threat of Entry*, 3 J. ECON. THEORY 306 (1971); Michael L. Katz & Carl Shapiro, *Network Externalities, Competition and Compatibility*, 75 AM. ECON. REV. 424 (1985); Paul Klemperer, *Competition When Consumers Have Switching Costs*, 62 REV. ECON. STUD. 515 (1995); Paul Klemperer, *The Competitiveness of Markets with Switching Costs*, 18 RAND J. ECON. 138 (1987); Paul Klemperer, *Entry Deterrence in Markets with Consumer Switching Costs*, 97 ECON. J. 99 (Supp. 1987); Paul Klemperer, *Markets with Consumer Switching Costs*, 102 Q.J. ECON. 375 (1987); Paul Klemperer, *Price Wars Caused by Switching Costs*, 56 REV. ECON. STUD. 405 (1989); Paul Milgrom & John Roberts, *Limit Pricing and Entry Under Incomplete Information*, 50 ECONOMETRICA 443 (1982); A. Michael Spence, *The Learning Curve and Competition*, 12 BELL J. ECON. 49 (1981).

77. A caveat is that infringing entry may be quite inefficient in the presence of learning curves and perhaps also if there are network externalities or consumer switching costs, because entry is likely to raise production costs or consumers' costs. *See* articles cited *supra* note 76; Paul Klemperer, *Welfare Effects of Entry into Markets with Switching Costs*, 37 J. INDUS. ECON. 159 (1988). But these are just special cases of the issue discussed *supra* in section II.C: from the social perspective these effects reduce the benefits of infringement *and* the costs of extending the patent life.

78. Patentees have at times been successful in claiming infringement damages for their own lost sales of noninfringing products, but no court to date has awarded damages because infringement transferred sales from the patentee to other noninfringing products. *See infra* section IV.B (discussing frontiers of patent damage litigation).

79. The stationarity intuition does not fail if the patentee must compete with a competitive fringe of noninfringing products. For example, if the patentee of a differentiated (and perhaps slightly better) mouse trap must vie in the marketplace against a competitive market

increase in the production of the noninfringing substitutes (and consequent reduction in the patentee's profit), then the patentee may be nontrivially harmed by the infringement. However, if limited infringement results in a decreased output of the noninfringing good,[80] the patentee would not be harmed by a little infringement. Furthermore, even if infringement does induce increased noninfringing output, the Ramsey intuition still typically applies: it is still likely to be socially preferred to allow a little infringement over a longer period (that is, award long-lived probabilistic patents) than to continue with the current regime of no infringement followed by free entry. Again, therefore, our results are only likely to be quantitatively rather than qualitatively affected.

While structural factors may independently constrain the pricing of many electronic and mechanical patents that must compete in more fluid high-tech environments, there is some evidence that monopoly pricing is much more of a concern with regard to pharmaceutical products. For example, F.M. Scherer has noted:

> From 1956 through the mid-1960s, the Pfizer Company and its four licensees sold the antibiotic tetracycline to druggists at a wholesale price of $30.60 per bottle of 100 capsules. Total sales at wholesale to drug stores exceeded $1 billion during this period. Production costs ranged between $1.60 and $3.80 per bottle; and when doubts about the validity of Pfizer's patent began to mount, several unlicensed firms began producing and selling tetracycline at approximately $2.50 per bottle wholesale. Many similar cases of price-cost margins on the order of 90 percent for patented drug products have been identified.[81]

---

of noninfringing traps, the patentee is likely to face a very elastic demand curve. The profit-maximizing price for this downward sloping, but relatively flat demand curve, is likely to be a much smaller percentage markup over the patentee's cost, but the logic of our main model still applies.

80. This will usually be the case in quantity competition, as our model assumes (and is the case for the demand and cost conditions in our simple model). More generally, the patentee will not be harmed by a little infringement when competition is between strategic substitutes in the terminology of Bulow, Geanakoplos, and Klemperer. *See* Jeremy I. Bulow et al., *Holding Idle Capacity to Deter Entry*, 95 ECON. J. 178 (1985); Bulow et al., *supra* note 74. In a price-competition model, however, strategic complements would be more likely. In that case, limited infringement might cause both the patentee and the noninfringing producers to lower their prices (and even perhaps increase their output) — so the patentee would be more likely to be significantly harmed by a small infringement. On the question of whether quantity competition or price competition is the more natural assumption, see Paul Klemperer & Margaret Meyer, *Price Competition vs. Quantity Competition: The Role of Uncertainty*, 17 RAND J. ECON. 618 (1986); Paul D. Klemperer & Margaret A. Meyer, *Supply Function Equilibria in Oligopoly Under Uncertainty*, 57 ECONOMETRICA 1243 (1989); and the articles cited therein.

81. SCHERER, *supra* note 69, at 450 (citing Henry Steele, *Monopoly and Competition in the Ethical Drugs Market*, 5 J.L. & ECON. 131 (1962); Henry Steele, *Patent Restrictions and Price Competition in the Ethical Drugs Industry*, 12 J. INDUS. ECON. 198 (1964)).

Such huge price-costs margins, combined with evidence of substantially lower prices offered by generic producers upon the expirations of the original patents,[82] suggests strongly that holders of pharmaceutical patents may be raising prices to points where the benefits of restricting market power are considerable. Certain types of uncertainty combined with delay might produce a net benefit with regard to pharmaceutical (and possibly to a lesser extent chemical) patents. Furthermore, as discussed above, pharmaceutical patents are more likely to have value at the end of twenty years and hence give lawmakers an opportunity to increase the duration of potential profitability. For pharmaceutical products, then, increasing patent duration while reducing certainty might maintain patentee expected profitability and thus maintain the incentive to innovate.[83]

III. Implications of the Theory

This Article does not propose that courts flip coins after fifty years to decide whether patents should be enforced. Yet appreciating the benefits of uncertainty and delay can inform many questions of policy at the margin. In particular, the model suggests that it is important to distinguish between two types of uncertainty in patent enforcement. The possibility that valid patents will be unenforced (Type I uncertainty) is more likely to be efficient than the possibility that invalid patents will be enforced (Type II uncertainty). Enforcing invalid patents creates ex post pricing distortions without enhancing innovation,[84] while our model showed that failing to en-

82. The patentee's markup can often be estimated by looking at the price generic producers are willing to charge once the patent expires. *See* Richard E. Caves et al., *Patent Expiration, Entry and Competition in the U.S. Pharmaceutical Industry*, *in* Brookings Papers on Economic Activity: Microeconomics 1, 36 (Martin Neil Baily & Clifford Winston eds., 1991) (finding that first generic entrant tends to charge 58.8% of patented price; with three generic entrants the generic price falls to 49.6% and with ten generic entrants the percentage falls to 29.4%); *Battling the High Cost of Drugs*, Harv. Health Letter (Harv. Med. Sch., Dept. of Continuing Educ.), July 1993, at 9 (noting that in Canada, generic manufacturers may copy patented drugs if they pay a royalty to the patent holder; the entry of the first generic competitor prompts an average price drop of 25-30%, and as others enter the price may drop up to 90%). There are, of course, other extreme examples. *See, e.g.*, *Moneyweek*, (CNN television broadcast, May 21, 1994) (reporting that prices on Naprosyn (by Syntex) fell 80% just 24 hours after the patent expired).

83. Also, as large corporations with multiple R&D projects and low-cost access to financial markets, pharmaceutical enterprises should be systematically less risk averse than smaller, single project inventors. Accordingly, risk aversion in the pharmaceutical context is less likely to contraindicate intentional uncertainty. *See supra* note 9 (discussing risk aversion).

84. This is especially true if courts enforce a patent that is not novel (whose idea is already embodied in the prior art). We note, however, that an optimal patent system might want to provide some lesser rewards for inventions that do not rise to the level of strict

force otherwise-valid patents could reduce the ex post distortions without reducing, or without substantially reducing, innovation incentives. This distinction suggests that policymakers should be more attuned to eliminating Type II uncertainty than Type I uncertainty.[85] The Type I uncertainty in the current system may not be as inefficient as previously thought — and reforms that move us toward a more certain regime with ex ante injunctions may not be cost effective.

A. *Reducing the Use of Preliminary Injunctions*

Since its creation in 1982, the Federal Circuit Court of Appeals has made it easier for patent holders to secure preliminary injunctions prohibiting competitors from producing potentially infringing products until the court has resolved the underlying dispute.[86] This strong preference for injunctive relief grows in part out of courts' essentialist view of patents as "property": "Patents must by law be given 'the attributes of personal property.' The right to exclude others is the essence of the human right called 'property.'"[87] The courts' protection of patent qua property becomes an end in itself, trumping all other conceptions of the good. For example, in a patent dispute in which Polaroid successfully sought to enjoin Kodak's manufacture of instant cameras, the court opined:

> Kodak's characterization of the public interest . . . misconstrues the very concept of public benefit. The public policy at issue in patent cases is the "protection of rights secured by valid patents." Courts grant — or refuse to stay — injunctions in order to safeguard that policy, even if those injunctions discommode business *and the consuming public.*[88]

---

patentability. Type II uncertainty might have the benefit of generating lower rewards for innovations that were, say, only quasi-novel.

85. General discussions of structuring legal rules to achieve the appropriate balance of Type I and Type II error can be found in Krier, *supra* note 3; A. Mitchell Polinsky & Steven Shavell, *Legal Error, Litigation, and the Incentive to Obey the Law*, 5 J.L. ECON. & ORG. 99 (1989); and Raaj Kumar Sah & Joseph E. Stiglitz, *The Architecture of Economic Systems: Hierarchies and Polyarchies*, 76 AM. ECON. REV. 716 (1986).

86. *See* Robert P. Merges, *Commercial Success and Patent Standards: Economic Perspectives on Innovation*, 76 CAL. L. REV. 803, 821 (1988); William A. Morrison, Note, *The Impact of the Creation of the Court of Appeals for the Federal Circuit on the Availability of Preliminary Injunctive Relief Against Patent Infringement*, 23 IND. L. REV. 169, 196 (1990) (finding post-Federal Circuit success rate for preliminary injunctions of 52%, which author found to be a statistically significant increase from the 36% rate from the preceding 29 years). *Epic Metals Corp. v. H.H. Robertson Co.*, 870 F.2d 1574 (Fed. Cir. 1989), is representative of this sea change.

87. Panduit Corp. v. Stahlin Bros. Fibre Works, 575 F.2d 1152, 1158 n.5 (6th Cir. 1978) (quoting 35 U.S.C. § 261).

88. Polaroid Corp. v. Eastman Kodak Co., 228 U.S.P.Q. 305, 344 (D. Mass. 1985) (emphasis added) (quoting Smith Intl., Inc. v. Hughes Tool Co., 718 F.2d 1573, 1581 (Fed. Cir.

The injunctive preference also grows out of a judicial misconception that denying injunctive relief (and making the potential infringer liable only for monetary damages) would cause unconstrained infringement. In 1987, the Federal Circuit found that absent injunctive relief, "patent rights will be flagrantly violated."[89] Especially for the judges of the Federal Circuit, the decision to deny injunctive relief is a decision to reduce substantially the expected profits of the patentee:

> The grant of a patent is the grant of the right . . . to exclude others . . . . Without this injunctive power of the courts, the right to exclude granted by the patent would be diminished, and the express purpose . . . to promote the progress of the useful arts, would be seriously undermined. . . . Without the right to obtain an injunction, the right to exclude granted to the patentee would have only a fraction of the value it was intended to have . . . .[90]

But our model has shown that even uncertain ex post enforcement will give rise only to *limited* amounts of infringement that need not substantially reduce the expected profitability of the patent. Potential infringers will see that as they compete down the market price they will be reducing the profitability of additional infringement and exacerbating the damages they will have to pay if the patent is ultimately enforced. Therefore, uncertain enforcement need not open a floodgate of infringement. Instead of taking an essentialist view that the "very nature" of property entails the right to exclude, we suggest that the nature of patents should entail offering sufficient rewards to stimulate innovation. Uncertain monetary damages — possibly combined with extended durations — are sufficient to achieve this end.

Doctrinally, the Federal Circuit has created a presumption that infringement will irreparably harm patentees. This presumption increases the risk that invalid patents will be enforced (Type II uncertainty). Our analysis suggests, however, that it would be wiser to shift toward a system with larger Type I uncertainty and smaller Type II uncertainty. *Flo-Con Systems, Inc. v. Leco Corp.*[91] provides an example of one potential shift in this direction. In *Flo-Con*, the court denied the patentee (Flo-Con) a preliminary injunction, and

---

1983)). *But see* Atari Corp. v. Sega of Am., 869 F. Supp. 783 (N.D. Cal. 1994) (denying preliminary injunction in part because 1200 permanent jobs would be lost).

89. H.H. Robertson, Co. v. United Steel Deck, Inc., 820 F.2d 384, 391 (Fed. Cir. 1987) (internal quotation marks omitted).

90. *Smith Intl., Inc.*, 718 F.2d at 1577-78.

91. 845 F. Supp. 1576 (S.D. Ga. 1993).

instead ordered the potential infringer (Leco) to put into escrow a litigation bond:

> A more workable solution than the blanket prohibition requested by Flo-Con is to require that Leco remit monthly payments to the registry of the Court similar to those it would make under a license agreement. Requiring the submission of such monthly payments assures compensation to Flo-Con if it prevails on the merits but avoids the problems associated with a blanket injunction — that is, it allows Leco to continue its presence in the marketplace and *preserves the benefits of competition.*[92]

Preliminary injunctions tend to resolve Type I uncertainty and prohibit interim infringement, leaving patent holders free to charge the monopoly price. Increasing the patent holders' burden to justify injunctions can promote competition and still preserve significant, but restricted, ex post protections.

In the pharmaceutical arena, the Drug Price Competition and Patent Term Restoration Act of 1984 (known as the Hatch-Waxman Amendments)[93] marks an analogous shift toward something very akin to preliminary injunctions. Hatch-Waxman allows the owner of a patented drug to stay for thirty months the FDA's approval of any generic application to produce the same drug — even if the generic's application claims that the initial patent is invalid or will not be infringed by the patented drug for which the applicant seeks approval.[94] Because a pharmaceutical manufacturer faces severe penalties for producing a drug without FDA approval, the thirty-month stay provides the patent owner with preliminary injunction protection. This aspect of the amendment is likely to reduce Type I uncertainty.

Moreover, parts of Hatch-Waxman have the effect of reducing the effective patent life. A safe-harbor provision in the Act allows generic manufacturers to engage in conduct "reasonably related" to obtaining regulatory approval for producing a version of a currently patented drug without infringing the patent.[95] This safe-harbor provision and other provisions shorten the effective duration of pharmaceutical patents by allowing generic manufacturers to file

92. 845 F. Supp. at 1583 (emphasis added).

93. Pub. L. No. 98-417, 98 Stat. 1585 (codified as amended in scattered sections of 21 U.S.C. and 35 U.S.C.).

94. *See* 21 U.S.C. § 355(j)(2)(A)(vii), (5)(B)(iii) (1994). This 30-month stay may be shortened by the court if the patent owner does not act reasonably or if the court finds the patent-in-suit is not infringed or is invalid. *See* 21 U.S.C. § 355(j)(5)(B)(iii).

95. *See* 35 U.S.C. § 271(e)(1) (1994); *see also* Brian D. Coggio & Francis D. Cerrito, *The Application of the Patent Laws to the Drug Approval Process* (Andrews Pharmaceutical, New York), Aug. 1997, at 12,421.

abbreviated applications for FDA approval, which facilitates the ability of competitors to begin selling generics soon after an original patent has expired instead of having to wait until the patent expires to prepare to seek FDA approval.[96]

Together these two aspects of Hatch-Waxman trade enhanced preliminary injunction protection in the short run for decreased effective patent duration. Our analysis suggests moving in just the opposite direction. Instead of using the requirement of FDA approval to eliminate interim infringement, we suggest that it might be better to allow limited interim infringement of generics and to compensate patentees with longer effective patent lengths.

B. *Vesting Decisionmaking Authority in Less Specialized Hands*

A major rationale for the creation of the Federal Circuit was to reduce uncertainty in the adjudication of appellate patent issues by vesting jurisdiction over all intermediate appeals in one specialized court. Commentators seem to agree that the court has been "successful" in reducing uncertainty — making patent litigation more predictable.[97] This is especially true because the Supreme Court rarely reverses Federal Circuit decisions — effectively making the Federal Circuit the centralized final arbiter of patent law. Moreover, the Supreme Court has recently affirmed a Federal Circuit decision that shifts decisionmaking power from juries to judges — by redefining questions of fact as questions of law.[98] Both of these reallocations of authority from nonspecialist judges and juries to the specialist Federal Circuit presumptively reduce the uncertainty of patent law — for the simple reason that specialists with greater knowledge and a thicker stock of precedent are more likely to rule consistently than nonspecialist decisionmakers who must first attempt to learn a rarely encountered and complicated body of statutory and common law. But as emphasized above, it is important to assess the effect of the Federal Circuit's increasing power on the relative mix of Type I and Type II uncertainty.

---

96. It should be emphasized, however, that other aspects of the Act were designed to increase the effective duration of pharmaceutical patents. *See* 35 U.S.C. § 156 (compensating patentees for the patent life that erodes during FDA review by adding to the patent term an amount of time roughly commensurate with the review period).

97. *See, e.g.*, Rochelle Cooper Dreyfuss, *The Federal Circuit: A Case Study in Specialized Courts*, 64 N.Y.U. L. REV. 1, 6-25 (1989).

98. *See, e.g.*, Markman v. Westview Instruments, Inc., 517 U.S. 370, 384-91 (1996) (holding that the interpretation of patent claim terms is the exclusive province of the court). *Markman* may also increase the use of interlocutory appeals to the Federal Circuit and thus — contrary to the themes of this Article — may speed the judicial award of preliminary and permanent injunctive relief.

Several commentators have emphasized that the decisions of the Federal Circuit are not just more predictable, but more pro-patent.[99] Scholars, for example, have estimated the Federal Circuit invalidates patents at substantially lower rates than did courts in the pre-Federal Circuit era.[100] This suggests that the Federal Circuit has not just moved to reduce uncertainty generally, but instead has worked to reduce Type I uncertainty much more than Type II uncertainty. Especially when combined with the strong preference for injunctive relief, this shift in enforcement runs against the implications of our model.

C. *Preferring Underinclusive Standards to Either Overinclusive Standards or Rules*

Our model suggests that underinclusive standards are likely to be more efficient than either rules or overinclusive standards. "Standards" (which can only be discovered ex post) are preferable to "rules" (which can be predicted ex ante) because standards create both the uncertainty and delayed ex post determination necessary to cause limited infringement.[101] And underinclusive standards are likely to create more Type I uncertainty (which we favor) than Type II uncertainty (which we disfavor).

For example, the four-part test for "nonobviousness"[102] may be the type of standard that creates both types of uncertainty — and which therefore has mixed efficiency qualities. But Professor Merges has argued that the commercial success doctrine — one of the four tests for nonobviousness — "grant[s] what amounts to a

99. *See* Merges, *supra* note 86, at 822-23; Morrison, *supra* note 86, at 187-88 (concluding that availability of preliminary injunctive relief against patent infringement has increased since the creation of the Federal Circuit).

100. *See* Merges, *supra* note 86, at 822 ("Between 1982 and 1985, the court invalidated only forty-four percent of the patents it adjudicated on appeal from trial courts, a marked contrast to the old invalidation rate of approximately sixty-six percent." (footnote omitted)); *cf.* Lawrence Baum, *The Federal Courts and Patent Validity: An Analysis of the Record*, 56 J. PAT. OFF. SOCY. 758, 760 (1974) (showing high rates of patent invalidation in courts before the creation of the Federal Circuit).

101. *See* Ian Ayres, *Preliminary Thoughts on Optimal Tailoring of Contractual Rules*, 3 S. CAL. INTERDISC. L.J. 1, 15-17 (1993); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559-60, 565-66 (1992).

102. To be valid a patent needs to be "nonobvious," a requirement that currently forces courts to consider (1) the scope and content of the prior art; (2) the level of ordinary skill in the art; (3) the differences between the claimed invention and the prior art; and (4) the secondary or objective evidence of nonobviousness. *See* Graham v. John Deere Co., 383 U.S. 1, 17-18 (1966); Michael J. Meurer, The Nonobviousness Standard and the Optimal Probability of Patent Validity (July 1995) (unpublished manuscript, on file with author). Having a patent's nonobviousness turn on the degree of ultimate commercial success may also induce ex ante uncertainty. *See* Merges, *supra* note 86, at 824-26.

quasi-presumption of patentability."[103] This position is consonant with the Federal Circuit's general pro-patent leanings, and it suggests that expanding the "nonobviousness" standard into new areas of enforcement is likely to be overinclusive. Rebuttable presumptions of patentability or preliminary injunctive relief are likely to decrease the risk of not enforcing valid patents (Type I uncertainty) and increase the risk of enforcing invalid patents (Type II uncertainty).

To see the disparate effects of over- and underinclusive standards on Type I and Type II uncertainty, it is particularly useful to contrast the effects of the "doctrine of equivalence" and the "reverse doctrine of equivalence." Under the doctrine of equivalence, courts may find infringement even though the accused product avoids the literal description of the patent claim.[104] But under the reverse doctrine of equivalence, courts may refuse to find infringement even though the accused product falls within the letter of the patent claim — so long as the accused product "has so far changed the principle of the device that the claims of the patent, literally construed, have ceased to represent his actual invention."[105] Expanding the current doctrine of equivalence is likely to be a movement toward overinclusion, and thus reduce the risk of not enforcing valid patents and increase the risk of enforcing invalid patents. In contrast, expanding the reverse doctrine of equivalence is likely to be a movement toward underinclusive enforcement and therefore have the opposite effects.[106] The thrust of our Article is that we can live with a standard-like expansion of the reverse doctrine of equivalence without seriously undermining the innovation incentives of patentees.

The choice of rules versus standards also impacts the relative mix of Type I and Type II uncertainty. For example, an important implication of the Patent and Trademark Office's (PTO's) substantive decision *not* to issue a patent is to eliminate an applicant's abil-

---

103. Merges, *supra* note 86, at 842.

104. *See* Warner-Jenkinson Co. v. Hilton Davis Chem. Co., 520 U.S. 17, 34-36 (1997); Alpex Computer Corp. v. Nintendo Co., 102 F.3d 1214, 1222 (Fed. Cir. 1996), *cert. denied*, 117 S. Ct. 2480 (1997); Valmont Indus., Inc. v. Reinke Mfg. Co., 983 F.2d 1039, 1041-44 (Fed. Cir. 1993).

105. Westinghouse v. Boyden Power Brake Co., 170 U.S. 537, 568 (1898).

106. Whether an expanded doctrine of equivalence actually leads to overinclusive enforcement turns on one's view about whether the preexisting (non-expanded) regime was over or underinclusive. However, even if the preexisting regime were underinclusive — so that Type I uncertainty is likely to be larger than Type II uncertainty — expanding the doctrine of equivalence is likely to have the deleterious uncertainty effect of decreasing Type I and increasing Type II uncertainty.

ity to subsequently sue others for patent infringement.[107] This is a "rule-like" effect of the current regime because it creates ex ante certainty that certain property is not protected. We do not favor underinclusive rules, because the certainty of Type I error would lead infringers to compete the market price all the way down to the competitive level — reducing the incentives to innovate. In contrast, underinclusive standards (where nonenforcement is only learned ex post) only induce limited amounts of infringement that (under the stationarity or Ramsey intuitions) need not reduce the innovation incentives.[108] On the margin, this militates against statutory or regulatory rules that are known ex ante and instead militates in favor of common law standards that often produce relatively delayed and uncertain adjudication — particularly if the common law is underinclusive.

D. *Allowing Patent Holders to Leverage Patent Market Power Into the Future*

While earlier sections have shown that consumers will generally have an incentive to trade some reduction in the patent's price for a longer patent,[109] the Supreme Court has prohibited such private

---

107. Unsuccessful applicants, however, do have limited rights of appeal. *See* 35 U.S.C. §§ 134, 141 (1994).

108. Imagine, for example, that the PTO has a 5% chance of committing both Type I and Type II error in reviewing applications — meaning that there is a 5% chance that valid applications would be denied and a 5% chance that invalid applications would issue. Under our current system there would still be some uncertainty whether the PTO's Type II error would be corrected (i.e., subsequent litigation might deny enforcement to the improvidently granted patents). Subsequent litigation might also create some Type I uncertainty — in that some correctly issued patents might be unenforced. But by giving deference to the PTO's initial decision, our current system reduces an important source of Type I uncertainty. Reducing the PTO's grounds for rejecting patent applications at the margin would reduce Type I error while increasing Type I uncertainty — both on the additional patents that would be issued, and on the patents that would in any case have been issued but for which the issuance of a patent is now a weaker signal of validity. In the extreme case, the patent office would simply become a registry that time-stamped patent claims to create a record for subsequent adjudication of validity.

The United States' "first to invent" standard may create both Type I and Type II uncertainty as compared to the "first to file" rule used by most other nations. While the Type I uncertainty would be valuable chiefly because it creates a limited class of infringers, the Type II uncertainty may be deleterious. *But see supra* note 84 (discussing possible benefits of Type II uncertainty). While "first to invent" disputes are currently resolved relatively quickly by PTO interference proceedings, our model suggests that it might be better to have delayed decisionmaking by (possibly less-specialized) courts so that we could sustain the benefits of limited infringement for a longer period.

109. *See supra* Introduction A and Section I.B. The parties could reduce distortion even more if the buyer paid a lump sum fee rather than a per-unit fee. However, sellers with patent market power may prefer royalties in order, for example, to implement second-degree price discrimination.

transactions. In *Brulotte v. Thys Co.*,[110] the Supreme Court struck down as unenforceable an agreement in which a consumer agreed to pay a royalty for more than twenty years on its purchases of a patented good. The Court's decision found especially problematic the fact that the royalty remained constant in the years after the patent had expired. But our analysis suggests that the Court's concern with leverage was misplaced. Negotiating a lower per-unit royalty in return for a longer royalty time period is likely to reduce the deadweight loss of supra-competitive pricing. And far from being suspect, the constancy of the royalty over time is prima facie consistent with the Ramsey intuition that the parties would have a joint interest in spreading the distortion evenly across time.[111] Our analysis thus suggests that *Brulotte* should be overruled.

IV. Alternative Implementations

The usefulness of limiting patentees' monopoly pricing power, however, does not turn on this particular method of achieving this end. Even if uncertainty and delay induce too many other inefficiencies, policymakers might explore a range of other policy instruments to reduce patent holders' monopoly power. This section first explores alternatives to expanded duration and then discusses two alternative mechanisms for restricting the market power of patentees.

A. *Expanding Geographic or Product Scope Instead of Duration*

The Ramsey intuition shows the utility of trading longer duration for reductions in patentee's market power. But the Kaplow ratio test suggests that other trades might also enhance efficiency.[112] It is generally efficient to expand a patentee's rights along the dimensions where the social cost per dollar of patentee profit is low and to contract a patentee's rights where the social costs per dollar of profit are high. The foregoing model showed that lengthening the duration right in return for a lesser pricing right was one such trade, but other trades are possible.

For example, an alternative to lengthening duration may be to widen the geographic scope of coverage. Instead of rewarding the

110. 379 U.S. 29 (1964).

111. *See* Meehan v. PPG Indus., 802 F.2d 881, 886 (7th Cir. 1986) (holding that when royalty payments extend unchanged beyond the life of a patent, the agreement is per se unlawful); Boggild v. Kenner Prods., 776 F.2d 1315, 1320-21 (6th Cir. 1985) (holding agreement per se unlawful when pre- and postexpiration royalties are the same).

112. *See supra* text accompanying notes 15-17 (discussing Kaplow).

patentee solely through a high tax on N countries, the Ramsey intuition suggests that it would be more efficient to produce an equal reward with a smaller tax on $N + 1$ countries. Thus, the United States might usefully spend greater effort to better enforce intellectual property rights in the rest of world rather than expending effort to increase the chance that valid patents will be enforced with certainty at home. Unsavory distributional effects of such efforts — particularly on less developed nations — might be compensated for in other ways.

Alternatively, it might be efficient to trade reduced patentee market power for patent protection of a wider class of products than might otherwise be covered.[113] Under certain conditions, broadening a patent's scope will be less costly (per dollar of patentee profit produced) than broadening a patent's duration.[114] And even when expanded duration is more efficient than expanded scope, it is still possible that expanding scope may be a more efficient way of generating patentee profits than expanding a patentee's power to raise price toward the monopoly level. When courts or lawmakers are unable to expand duration, expanding patent breadth may be a useful way to compensate for reductions in uncertainty.

B. *Partial Damages*

Giving the nonpatent holders the option of producing for a fee that is less than the monopoly markup would induce limited amounts of interim production.[115] The probabilistic patent model could itself be reinterpreted as a system in which w was interpreted not as a probability of enforcement but as the proportion of "make

113. *See* Gilbert & Shapiro, *supra* note 2, at 109 (discussing this possibility); Klemperer, *supra* note 2, at 121-23 (discussing in a more general model when this is desirable).

114. Klemperer analyzes the situations in which it is more desirable for society to reward a patentee through a broader scope of coverage rather than by a longer duration. *See* Klemperer, *supra* note 2, at 121-23. If consumers would have relatively similar valuations for the patented product when no close substitutes were available, but would have relatively dissimilar willingness to pay for the patented product if low-price close substitutes were available, then it is relatively more desirable to give the patentee control over the close substitutes: that is, the patent should be broadened so that the close substitutes would infringe. If the opposite is true — for example, if potential consumers have varying levels of need for a product (because they would use it with different frequencies) but have similar strengths of preferences between an easy-to-learn product and harder-to-learn copies — then a longer-lived but more narrowly construed patent covering just the "ideal" easy-to-learn product is socially preferred.

115. The potential utility of compulsory licensing is discussed in Merges, *Of Property Rules, supra* note 23; Ralph Oman, *The Compulsory License Redux: Will It Survive in a Changing Marketplace?* 5 CARDOZO ARTS & ENT. L.J. 37 (1986); and J.H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 COLUM. L. REV. 2432 (1994).

patentee whole" damages that infringers would have to pay. In this model, interim producers would be certain to pay damages, but because the damages would not increase the patentee's payoffs to the monopoly level, limited amounts of infringement would occur.[116] However, unlike compulsory licenses, which set the patent royalty to yield a reasonable return above cost,[117] this partial damages regime would be set to give the patentee damages equaling a large proportion (on the order of say 70-90%) of the losses relative to full monopoly profit. And unlike current compulsory license regimes, the partial damages regime could estimate the award either ex post or concurrently as production proceeds.[118] The information available to the government actor ex post would be much greater than with traditional compulsory licenses where the royalty is determined before production begins.

In order for courts to implement a shift toward partial make-whole damages, they would first need to reject the strong preference for injunctions — especially preliminary injunctions. The essentialist notion of "patents" as "property" that by nature entails the right to exclude has led courts to discourage interim infringement even when the consumer costs of monopoly dramatically outweigh the benefits to the patentee (and any plausible impact on innovation incentives). If courts can free themselves of this notion, however, they can shift away from issuing injunctions, as the issu-

---

116. Protecting the patentee's entitlement with partial "liability" damages may not endogenously induce the patentee to delay judgement. *See generally supra* section I.D (discussing factors influencing delay). If damages do not make the patentee whole, the patentee may prefer to gain an early injunctive decision that would give the patentee monopoly profits for the remainder of the patent's life. (If, however, winning judgment only gave the patentee a right to partial prospective damages, then the patentee would prefer delay as before.) Algebraically, if k is defined to be the proportion of make-whole damages that the patentee would receive with probability w, then the patentee's expected profits from not receiving judgment until the end of the patent life would only be greater than the expected profits from receiving probabilistic injunctive judgement immediately if:

$$q_p p + kw\,(\pi_M - q_p p) > (1 - w)\,0 + w\pi^M$$

In the linear demand example,

$$q_p = p = \frac{50kw}{(2 - kw)}$$

so delaying judgment will only be the patentee's preferred strategy if:

$$w < \left(\frac{kw}{2 - kw}\right)^2 \left(\frac{1 - kw}{1 - k}\right)$$

117. *See* F.M. SCHERER, THE ECONOMIC EFFECTS OF COMPULSORY LICENSING 43 (1977).

118. If desired, there could also be a final settling up based on ex post information. Of course, courts could demand that infringers post interim bonds to make sure they have the wherewithal to ultimately pay the ex post award.

ance is not mandated by statute but is within the courts' equitable discretion.

If courts move away from injunctive remedies, there are several elements of the current damages calculations that could be transformed into what in effect would be a partial damages regime. Unlike trademark and copyright law, patent law does not calculate damages so as to force infringers to disgorge any benefits of infringement.[119] The patent statute only requires that patentees be paid make-whole damages — that is, "damages adequate to compensate for the infringement."[120] While true make-whole damages will still often represent what Calabresi and Melamed termed a "property rule," which will deter interim infringement, the methods of proving and calculating such damages can effectively transform them into a partial damages regime. For example, the widely used standard of *Panduit Corp. v. Stahlin Bros. Fibre Works, Inc.*[121] creates a standard of proof that is likely to facilitate interim infringement:

> To obtain as damages the profits on sales he would have made absent the infringement, i.e., the sales made by the infringer, a patent owner must prove: (1) demand for the patented product, (2) absence of acceptable noninfringing substitutes, (3) his manufacturing and marketing capability to exploit the demand, and (4) the amount of the profit he would have made. . . .
>
> . . . .
>
> When actual damages, e.g., lost profits, cannot be proved, the patent owner is entitled to a reasonable royalty. A reasonable royalty is an amount "which a person, desiring to manufacture and sell a patented article, as a business proposition, would be willing to pay as a royalty and yet be able to make and sell the patented article, in the market, at a reasonable profit."[122]

The difficulties that patentees frequently have in proving the four *Panduit* prerequisites often mean that instead of being awarded lost profits (what amounts to make-whole damages), patentees must settle for the smaller reasonable royalty measure. Our analysis suggests that this is generally socially beneficial. There are several

119. Trademark and copyright infringers are required to pay the greater of (1) what the rights holder lost from infringement, or (2) what the infringer gained. *See* 17 U.S.C. § 504(b) (1994) (copyright); 15 U.S.C. § 1117(a) (1994) (trademark). In our simple model, infringers in equilibrium disgorge all of their expected profits to the patentee. But in richer models, (for example, if infringers had lower costs than the patentee) not all profits from infringement would be disgorged.

120. 35 U.S.C. § 284 (1994).

121. 575 F.2d 1152 (6th Cir. 1978).

122. 575 F.2d at 1156-57 (quoting Goodyear Tire and Rubber Co. v. Overman Cushion Tire Co., 95 F.2d 978, 984 (6th Cir. 1937) (internal quotation marks and citations omitted)).

other frontiers[123] of the battle for lost-profit damages — including issues of price erosion;[124] the market share rule;[125] lost sales on unpatented products;[126] and postexpiration sales[127] — where our analysis suggests that awarding lower "partial" damages might have the salutary effect of inducing limited amounts of infringement without unduly lessening innovation incentives.

C. *Duopoly Auction*

A system that required patentees to auction the right to use the innovation could also induce interim infringement. A patent would give the holder two entitlements: the right to be one of only two producers of the product,[128] and the right to receive the proceeds from the auction selecting the second producer of the product. For concreteness, a year after receiving a patent, the government might auction the right to use the innovation to one additional firm. In a simple model, competition among firms to be the second producer would tend to raise the auction price sufficiently so as to give the patentee all of the second producer's expected profit.[129] Just as in

123. *See* ROBERT P. MERGES ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGY AGE 306 (1997) (notes on the Frontiers of Lost Profit Damages).

124. Patentees at times convince courts that they have lost profits not only because they sold fewer items, but also because infringement erodes the market price and therefore the profits on the items they continue to sell. *See, e.g.*, Lam, Inc. v. Johns-Manville Corp., 718 F.2d 1056 (Fed. Cir. 1983). We generally favor awarding additional damages for price erosion so as ultimately to limit the amount of infringement, and we would even favor a price erosion adjustment to the calculation of reasonable royalty. But the details of calculating the rate at which the erosion damages accrue must be sufficiently conservative so as not to deter all infringement. Judge Frank Easterbrook, sitting by designation as a district court judge, has laid out the most sophisticated analysis of how to go about calculating price erosion damages. *See In re Mahurkar Double Lumen Hemodialysis Catheter Patent Litig.*, 831 F. Supp. 1354, 1385-93 (N.D. Ill. 1993).

125. The *Panduit* test requires the absence of noninfringing substitutes. Some courts, however, have calculated lost profit damages where at least one seller in a market is a noninfringing competitor by assuming that the patentee's market share relative to the noninfringer would have remained the same in the absence of infringement. *See* State Indus. v. Mor-Flo Indus., 883 F.2d 1573, 1579 (Fed. Cir. 1989).

126. Patentees — under the "entire market value rule" — have recently been claiming damages for lost profits on unpatented components that they claim would have been made but for infringement of a related patented product. *See* TWM Mfg. Co. v. Dura Corp., 789 F.2d 895, 900-01 (Fed. Cir. 1986).

127. Patentees have at times gained additional damages by arguing that infringement during the term of the patent gave the infringer a head start on postexpiration sales. *See* TP Orthodontics, Inc. v. Professional Positioners, Inc., 17 U.S.P.Q.2d 1497, 1504-06 (E.D. Wis. 1990).

128. The current regime gives coinventors just this right in one situation: coinventors, as tenants-in-common, each own an undivided one-half interest in the patent. *See* Drake v. Hall, 220 F. 905, 906 (7th Cir. 1915). Thus, they each have a right to practice the invention and to exclude anyone except their fellow inventor from practicing the same.

129. For richer models of auctions, see THE ECONOMIC THEORY OF AUCTIONS (Paul Klemperer ed., 1999), and Paul Klemperer, *Auction Theory: A Guide to the Literature*, 13 J.

the probabilistic regime, the patentee would earn all of the industry profits during the life of the patent: it would earn profits from its own sales directly, and profits from the other producer's sales through the lump sum payment that this second producer paid in the auction. But the market price in this duopoly would likely be lower than the monopoly price the patentee would charge under the current regime. Consumers may be better off paying duopoly prices for, say, twenty-three years than paying monopoly prices for twenty.[130]

CONCLUSION

Allowing patentees to profit from their patents furthers the valuable social objective of encouraging innovation. But as Louis Kaplow has already observed: "It is simply not true that all activities generating equal profits impose equal damages upon society."[131] Enlightened intellectual property law will try to allow patentees to profit in ways that impose the smallest cost on society.[132] This Article has pointed out that the profits coming from the last increment of monopoly pricing impose disproportionate costs on society (relative to, for example, allowing the patentee to earn restricted periodic profits for a longer period of time). The traditional core method of rewarding innovation — allowing patent holders to charge a monopoly price — does not pass Kaplow's cost-effectiveness test because patentees do not extract their reward in the least costly way.

This Article has also shown that there are policy options to constrain patentees' pricing power that do not assume unrealistic government information. In particular, uncertainty and delay in patent

---

ECON. SURV. (forthcoming May 1999). For example, the problems discussed in Paul Klemperer, *Auction with Almost Common Values: The 'Wallet Game' and Its Applications*, 42 EUR. ECON. REV. 757 (1998) may be important.

130. This oligopoly auction idea is similar to the "permissive" patent proposal of La Manna et al., *supra* note 44, in that both would allow multiple producers for each innovation. But the purpose of permissive patents would be to reward nonplagiarist copiers for their attempts at innovation, while the purpose of the oligopoly auction is merely to restrict the exploitation of patentee power when that exploitation is not cost effective.

Michael Kremer, *supra* note 54, 1146-48, suggests that the government should auction the monopoly right to the patent, but with some predetermined probability cancel the result of the auction ex post and offer to pay the patent holder the final price determined by the auction "times some constant markup which would reflect the typical ratio of social to private value," Kremer, *supra* note 54, at 1146, in return for allowing free production of the innovation by all. Of course, this approach would require substantial government funds.

131. Kaplow, *supra* note 15, at 1821.

132. While this Article has focused on patent law, the stationarity and Ramsey intuitions might also usefully inform copyright law (and possibly even trademark law).

litigation may have the unforeseen benefit of restricting patentees' market power by inducing limited amounts of interim infringement. Because the infringing firms probabilistically compensate the patentees (and because of the stationarity intuition), the patentees' reduced selling prices do not need to unduly reduce their expected profits. Furthermore, the government has sufficient information to trade-off more substantial reductions in the price of the patented good for an offsetting increase in the patent life. These insights about uncertainty and delay illuminate a number of policy questions, especially those concerning the relative advantages of preliminary injunctions versus less expansive monetary damages.