Progress in the Useful Arts: Foundations of Intellectual Property Law in Growth Theory

by

Robert Cooter
Herman Selvin Professor of Law, Berkeley Law School
rcooter@law.berkeley.edu

This paper is based on the draft of a book manuscript entitled The Falcon's Gyre: Law and Economics of Innovation. Download the latest version by Googling "cooter falcon gyre" or by clicking http://scholarship.law.berkeley.edu/books/1/.

Revised 14 October 2015. I benefitted from comments at previous presentations at the Center for Technology, Innovation & Competition, University of Pennsylvania Law School; the law and economics seminars at Stanford Law School, the University of Virginal Law School, the University of Arizona Law School; and the law faculty seminar at Florida State University and Washington University in St. Louis.

Abstract:

The US constitution authorizes Congress to enact intellectual property (IP) law to promote progress in the useful arts. When innovation is rapid, the welfare gains from growth overtake the welfare losses from static inefficiency (the *overtaking principle*). Consequently, maximizing welfare requires maximizing innovation, not balancing innovation and static efficiency. When combined with the standard analysis of market power, the overtaking principle simplifies the analysis of IP. Market power increases prices, but sellers gain less than buyers lose (deadweight loss). The resulting welfare effects depend on the product's use. When innovations are bought for non-innovative activities, wealth transfers from the latter to the former, which increases the growth rate. With rapid innovation, the increase in the growth rate overtakes losses from static inefficiency, so welfare increases. In contrast, when innovations are bought for innovation, wealth transfer among innovators. The deadweight loss decreases the overall profitability of innovating and reduces welfare. This analysis implies the *separation principle of intellectual property:* IP protection should be strong against using an innovation for consuming and producing, and IP protection should be weak against using an innovation to innovate. The prescription of weak IP rights against innovating, however, has an important qualification. "Fertility" refers to an innovation's power to stimulate other innovations. The *fertility principle* asserts that maximizing innovation requires strengthening IP for fertile innovations against less fertile innovations, until the increase in growth from more fertility equals the decrease in growth from more deadweight loss.

Robert Cooter[*]

"Progress in the Useful Arts" -- Foundations of Intellectual Property Law in Growth Theory[**]

> "The Congress shall have power…To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries…" -- Article I Section 8 of the U.S. Constitution.

The US constitution authorizes Congress to enact intellectual property (IP) law for a stated purpose: progress in the useful arts. The usefulness of the arts can be measured by economic indexes such as gross national product in national accounting, net benefits in cost-benefit analysis, and the quality of life in non-traditional measures of wealth. Although controversial, growth in these indexes caused by innovation provides the best overall measure of progress in the useful arts.

Economics distinguishes static and dynamic causes of increases in economic value. Static efficiency is achieved by maximizing the value of given goods produced by given techniques. "Given" means "the same", or "constant". Static efficiency increases by improving the mix of goods produced, or better "resource allocation". The model of perfect competition formulates ideal conditions for static efficiency. To approximate these conditions, antitrust law usually suppresses market power.

Alternatively, economic innovation discovers better goods to make and better techniques to making them. Innovation builds on itself, so it causes sustained growth. Growth connects to IP through the fact that the rate of

[*] Robert Cooter is Herman Selvin Professor of Law, University of California at Berkeley.

[**] I am grateful for comments from Aaron Edlin, Elhanan Helpman, participants in the Stanford Law and Economics Seminar, including Mitch Polinsky, Mark Lemley, Marus Cole, Michael Klausner, and George Triantis; also the University of Virginia Law and Economics Seminar, including Michel Barzuza, Curtis Quinn, John Dufy, and Dotan Oliar.

innovation usually increases with the reward for innovating. To increase the reward for innovating, IP law grants market power to innovators.

As explained, IP mostly enhances market power for the sake of innovation, and antitrust mostly suppresses market power for the sake of efficient production. When the two purposes collide, law and policy must balance dynamic and static effects on the welfare of people. They balance when a small increase in market power for innovations increases welfare by the same amount as the resulting loss from distorting resource allocation. This is the "balancing test" for optimal market power. The optimum favors strong market power for innovators in dynamic industries and weak market power for producers in static industries. Thus lawmakers should ideally strengthen intellectual property rights until the increase in welfare from faster innovation equals the decrease in welfare from static inefficiency.

Economists who study IP law usually think that the law maximizes social welfare if it strikes this balance. According to this Article, the balancing test is unnecessarily complicated for purposes of intellectual property law. Instead of the balancing test, the next section argues that IP law maximizes social welfare when it maximizes innovation. IP law can dispense with balancing because, according to the "welfare overtaking principle", the welfare gains from rapid innovation overtake the welfare gains from static improvements in efficiency or distribution. Instead of balancing dynamic and static effects, the overtaking principle gives all weight to dynamic effects and no weight to static effects. Compared to balancing, maximizing innovation simplifies the interpretation of IP law and provides a better account of IP's constitutional purpose.

Welfare Overtaking

Robert Lucas, the economist who won the Nobel prize in 1995, famously commented, "Once one starts to think about economic growth, it is hard to think about anything else." Compared to sustained growth, other sources of national wealth are insignificant. Compounded over a century, 2% annual growth (roughly the growth rate of the US economy over a century) increases wealth

more than 7 times, and 10% annual growth (roughly the growth rate of the Chinese economy from 1980 to 2010) increases wealth by almost 14,000 times.[1]

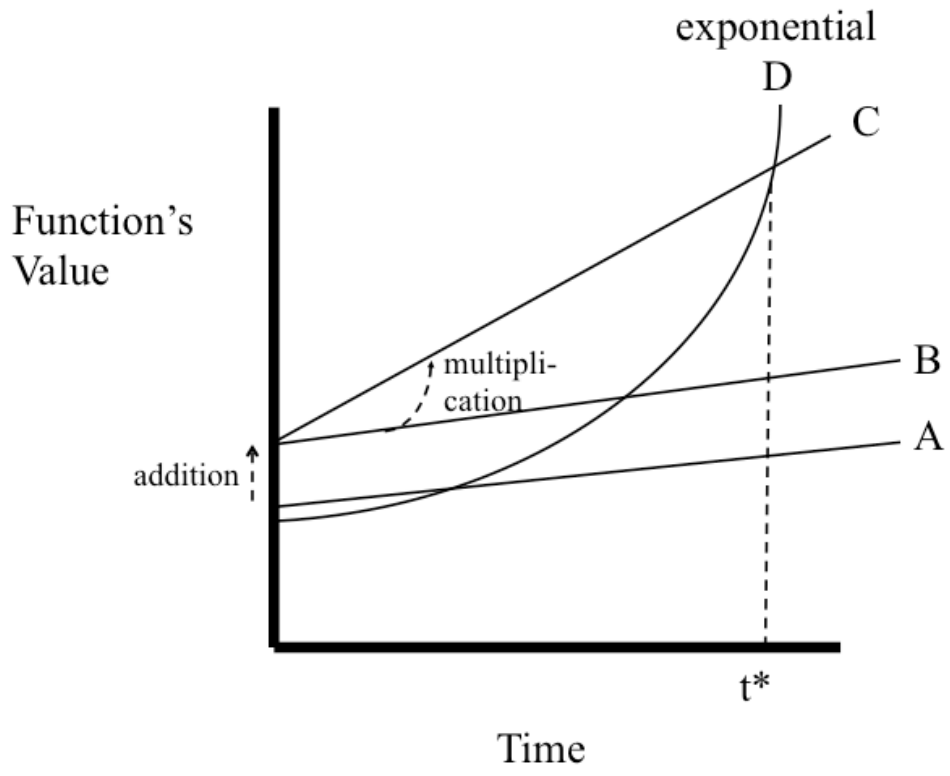
People under-estimate the effects of compound growth because they confuse it with absolute growth. They forget that a 5% increase is much larger for a teenager than a toddler. Behind these observations is a mathematical truth: an economy that increases at a constant proportional rate will overtake an economy that increases at a constant absolute rate. Figure 1 depicts this fact. The vertical axis in Figure 1 represents a function's value, and the horizontal axis represents time. Start with function A, whose value increases at a constant absolute rate with time, as indicated by A's constant slope. An addition to the value of A at time 0 shifts A up and yields B. B starts at a higher level than A, and B grows at the same absolute rate as A. The multiplicative increase to C results in growth at a higher absolute rate for C than B. Now contrast C to D that increases at a constant proportional rate. D starts below C at time 0 and overtakes C at time t*. So a function with constant proportional growth rate will overtake a function with constant absolute growth. Similarly, a function with higher proportional growth will overtake a function with lower proportional growth.

[1] Here's a table of size reached by an economy that starts at 1 and grows at various rates and years.

Growth rate	1 year	5 years	10 years	25 years	50 years	100 years
0%	1.00	1.00	1.00	1.00	1.00	1.00
1/2%	1.005	1.01	1.05	1.13	1.28	1.65
1%	1.01	1.05	1.10	1.28	1.64	2.70
2%	1.02	1.10	1.22	1.64	2.69	7.24
5%	1.05	1.28	1.63	3.39	11.47	131.50
10%	1.10	1.61	2.59	10.83	117.39	13,780.61

*size of economy $= (1+r/100)^t$, where r=percentage growth rate, and t=years of growth.

Figure 1. Exponential Growth Overtakes Additive and Multiplicative Increases



The mathematics of overtaking in Figure 1 is unaffected by the function's interpretation. The function might represent income per capita. Under this interpretation, Figure 1 shows that an exponential increase in income per capita overtakes an additive or multiplicative increase. Furthermore, a faster exponential increase in income overtakes a slower increase. This is "income overtaking." Alternatively, the function could represent social welfare. The same propositions apply to "welfare overtaking."

The essential difference between static and dynamic processes is how they build on themselves. Static processes cause additive and multiplicative changes. In Figure 1, the shift from A to B represents one-time growth, not growth that builds on itself. The shift from B to C represents growth that builds on itself at a constant rate. More efficient resource allocation resembles shifts

from A to B, or B to C. In contrast, the shift to D represents growth that builds on itself at a constant proportional rate. Increases in the rate of innovation resembles a shift to D. The relatively static improvements in A, B, and C seem unimportant relative to the dynamic improvement in D.

Few people regard income as an end in itself. Instead, most economists and some philosophers regard income as an instrument for welfare. In this view, welfare is the final measure of economic value for purposes of law and policy. Almost everyone can agree about some dramatic effects of income on welfare. Thus life expectancy at birth is 83 years in Japan and 66 years in Bangladesh.[2] Enrollment in secondary school is 98 percent among Japanese children of the appropriate age and 42 percent in Bangladesh.[3] Facts like these make almost everyone agree that welfare is much higher in Japan than Bangladesh, primarily because Japan is much wealthier than Bangladesh.

With exponential growth, what we thought were ends turn out to be means. Static efficiency is a case in point. Efficient resource allocation improves welfare directly. However, this direct effect is unimportant compared to growth. Given exponential growth, resource allocation affects welfare most through its effects on innovation and growth.

In these respects, static redistribution of income resembles static efficiency. The amount of welfare created by given income depends on who gets it and how it is spent. An old tradition in economics holds that more money benefits the poor more than the rich.[4] The "marginal utility of money" is higher for the poor because they use money to satisfy more urgent needs than the rich, as when the poor buy bread with an additional dollar and the rich buy cake. Consequently, redistribution from rich to poor increases welfare directly.

---

[2] Life expectancy at birth, total (years)) in 2008, World Bank. http://data.worldbank.org/indicator/SP.DYN.LE00.IN. Note that life expectancy is in the 40s in many African countries.

[3] Data & Statistics on Education for 2007 and 2008, World Bank, http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTEDUCATION/0,,contentMDK:20573961~menuPK:282404~pagePK:148956~piPK:216618~theSitePK:282386,00.html.

[4] In Adler's theory (see preceding footnote), social welfare increases with individual wealth, and social welfare increases with a more equal distribution of wealth or utility across individuals.

However, the indirect effects through growth are more important. Like efficient resource allocation, equality mostly contributes to welfare as a means to growth, not an end. Redistribution that increases growth increases welfare, including the welfare of the poor. To illustrate historically, China apparently destroyed growth in the Cultural Revolution, which commenced in the 1960s and expired by 1975, because strict equality undermined the incentive to innovate. Reversing after 1980, China allowed innovators to keep much of the value that they created. Equality decreased, economic growth exploded, and poverty plummeted. The lowest wage earners in China benefited more from a faster growth rate in national income after 1980 than from more equality in the 1970s. [5]

Conversely, too much inequality can slow growth, especially by undermining the health and education of workers. Thus improved education and health of workers would assuredly increase their creativity in, say, the U.S.A. or the Philippines. According to the overtaking principle, however, redistributing from rich to poor is unimportant in itself. Inequality is optimal when it maximizes growth of income and welfare, including the income and welfare of the poor. For the welfare of the poor, the important question about equality is whether it increases or decreases the growth in the incomes of the poor.

The increase in social welfare from faster exponential growth in income usually overtakes additive or multiplicative increases in income, including more static efficiency and more equal redistribution. I call this proposition the "welfare overtaking principle". Once you start thinking about welfare overtaking, it's hard to care about static efficiency and income redistribution for their own sake.

The welfare overtaking principle makes various assumptions about discounting future utility and the marginal utility of income that I will not discuss

[5] David Mktrian and I have compared the gain in welfare from actual growth and hypothetical redistribution in China and the USA in recent years. To illustrate our preliminary findings, assume hypothetically that income could be redistributed without loss in China so that each quintile received an equal income. If welfare is a logarithmic function of income, then this hypothetical redistribution from rich to poor would substantially increases welfare in the poorest quintile of the population. However, it would only take three years of growth at recent rates for the welfare of the poorest quintile to overtake the level achieved by this hypothetical redistribution.

here.[6] This article concerns intellectual property, not philosophy or economic theory. This is not the place to explore theoretical puzzles and political disagreements about the relationship between income and welfare. The welfare overtaking principle permits us to circumvent these disputes, so long as we limit our discussion to economies with exponential growth in income. If people agree that exponential growth in income is causing exponential growth in welfare, then they must agree that welfare overtaking will occur (although they may disagree over how soon it will occur). Our theory of IP law appeals most when applied to economies where exponential growth in income causes exponential growth in welfare.

The constitution authorizes Congress to create IP law to promote progress in the useful arts, which I interpret as increasing economic growth through innovation. Given exponential growth, the overtaking principle justifies the claim that intellectual property law will increase national welfare more by maximizing innovation than by balancing growth and static efficiency. Next I will use the overtaking principle to establish the foundations of intellectual property law in growth economics.

I. Ventures

Innovation especially occurs through business ventures. A bold ship's captain in seventeenth-century England proposes to investors that they finance a voyage to Asia for spices.[7] The voyage is costly and risky, but if it succeeds, the spices will be worth a fortune. Seventeenth-century spice voyages involved up-

[6] Assume that utility in a generation is a function of its consumption, and social welfare equals a discounted sum of the utilities of the generations. Consider two sequences of consumption in an infinite number of generations. In the first sequence, initial consumption is higher and growth is lower. In the second sequence, initial consumption is lower and growth is higher. If the social welfare function is reasonably convex in the utility of the generations, and the utility function of each generation is reasonably convex in consumption, then the sum of utilities in the second sequence eventually overtakes the sum of utilities in the first sequence. A theorem along these lines is proved in Robert Cooter & Aaron Edlin, *Maximizing Growth vs. Static Efficiency or Redistribution*, BERKELEY LAW AND ECONOMICS WORKING PAPER (2011). I am now constructing simulations of various paths of welfare depending on growth and distribution of income.

[7] This characterization of the spice trade is based on Ron Harris, "Law, Finance and the First Corporations", in James J. Heckman, Robert L. Nelson & Lee Cabatingan, eds., Global Perspectives on the Rule of Law ((2009, Routledge).

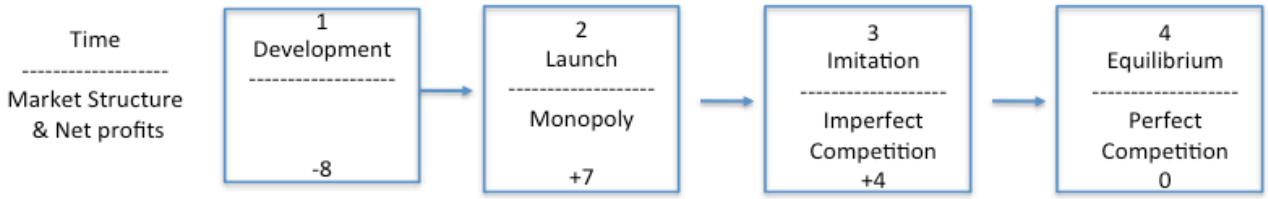
front investment, high risk, and high return. Similarly, an engineer in Silicon Valley in 2010 has an idea for a new computer chip. Development is costly and risky, but if it succeeds, it will be worth a fortune. Twenty first century technological innovations involve up-front investment, high risk, and high return.

A profitable business venture often has a life cycle like Figure 2. The venture begins with the development of a new idea in period 1, which costs 8. By developing the idea into a product, the innovator acquires a valuable secret or patent, or perhaps a cluster of secrets or a portfolio of patents. After development, the innovation is launched and marketed to buyers. When launched in period 2, the innovation has no competitors, so the innovator is a monopolist who receives a payoff of 7. In period 3, imitators develop competing products that substitute for the innovation, which reduces the innovator's payoff to 4. In period 4, imitations improve and competition intensifies. Taking competition to its logical extreme, the imitations become perfect in period 4, so the market is perfectly competitive and the innovator's payoff is zero.[8] A successful business venture earns extraordinary profits when launched and imitated, and ordinary profits when perfectly substituted. Summing over the life cycle in Figure 2, the venture's net payoff equals +3.

Production of the innovation will continue beyond time 4 under conditions approximating perfect competition until the product becomes obsolete. The product becomes obsolete when a new innovation destroys the old one's value and the cycle begins anew. When ventures like the one in Figure 2 repeat themselves, one innovation follows another, and the path of net social benefits spirals up like the falcon's gyre.

[8] Perfect competition drives the prices of all goods to their cost of production. Profits are zero after including the cost of capital in the other costs of production. The cost of capital equals the ordinary rate of profit in alternative uses.

Figure 2. Life Cycle of a Profitable Venture



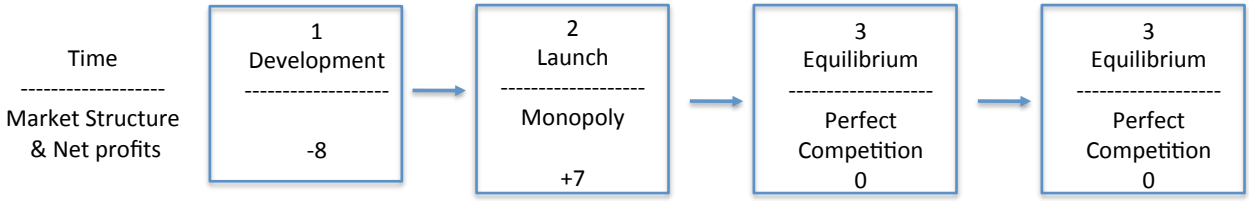
The first firm to develop and market an innovation often enjoys market power temporarily. Precedence conveys advantage three ways. First, precedence may enable the innovator to get big first and enjoy increasing returns to scale – first in time is first in might. Second, secrecy often extends this advantage by slowing the discovery's dissemination to competitors. Third, instead of secrecy, patents and copyright can create market power by legal fiat – first in time is first in right. These three sources of market power for innovators – scale economies, secrecy, and legal fiat -- especially relate to three bodies of law -- antitrust, trade secrets, and intellectual property. This paper concerns the third source of market power for innovators -- intellectual property law.

Figure 2 depicts a profitable venture, but most ventures fail and lose money. Recent U.S. data suggests that 40% of new businesses survive and 60% disappear within four years.[9] Figure 3 depicts a losing venture. The innovator in Figure 3 spends 8 in period 1 to develop the product. Many innovations fail before completing development and beginning production, without recouping any development costs. The innovator in Figure 3, however, is a little more successful and brings the product to market. When the innovation is launched in period 2, the innovator has no competitors and enjoys profits of 7. The only difference between Figures 2 and Figure 3 is in period 3. In Figure 2 the

[9]44% of U.S. businesses started in the 1990s still existed 4 years afterwards. See Knap, A. E. (2005). "Survival and longevity in the Business Employment Dynamics data." Bureau of Labor Statistics of the U.S. Department of Labor. The number given by Dun and Bradstreet is 37% as cited in "Some of the Reasons Why Business Fail and How to Avoid Them," Entrepreneur Weekly, Issue 36, 3-10-96.

innovation yields a payoff of 4 in period 3 because it is better than the imitations. In Figure 3, however, the imitations in period 3 are just as good as the original – they are perfect substitutes. Perhaps the innovator cannot keep a secret or it has no patent. In Figure 3, the market reaches a competitive equilibrium in period 3, so the innovator's payoff is 0 in period 3, as well as in period 4. Summing over the life cycle, the venture's net payoff equals -1. The innovator in Figure 3 cannot recover the cost of developing the new idea.

Figure 3. Life-Cycle of an Unprofitable Venture



The innovator sometimes has the advantage as in Figure 2, and the innovative venture is profitable. The imitator sometimes has the advantage as in Figure 3, and the innovative venture is unprofitable. The difference between a profitable venture in Figure 2 and an unprofitable venture in Figure 3 is the ease of imitation.

To slow imitators, innovators often try to keep their innovations secret,[10] and the law of trade secrets assists them.[11] Instead of secrets, another route to extraordinary profits is patenting, which requires disclosing the invention to the public. The creator of a patentable invention must decide whether secrecy or a patent is more profitable, and proceed accordingly.

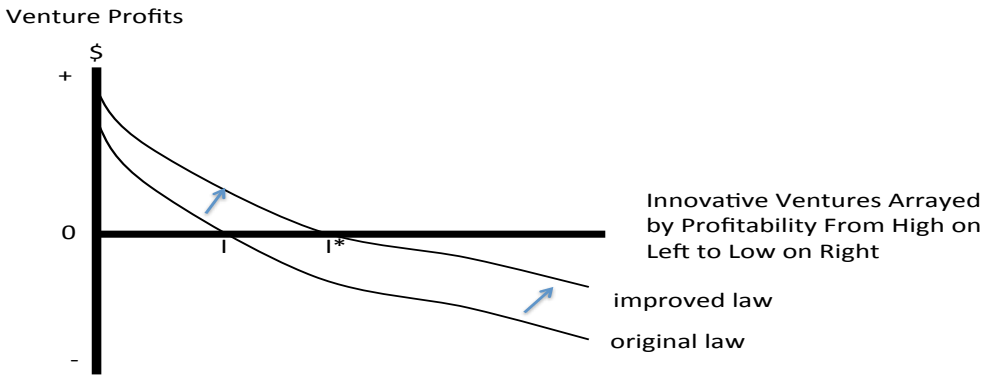
To understand the rate of innovation, imagine an array of new ideas that differ according to the expected profitability of developing them. The expectation

[10] Some innovations reduce to explicit information that is easy to copy, like a recipe or a computer code. Being easily copied, explicit information is intrinsically hard to keep secret. Unlike a computer code, other innovations involve implicit information that is irreducible to simple communication. Implicit information is often imbedded in a practice or organization, like judgment in mixing chemicals, art in baking cakes, or methods to motivate salesmen. Information is implicit when someone knows how to do something that is hard to communicate. Because communication is hard, implicit information is easier to keep secret than explicit information. To steal another company's implicit information, you need to hire its employees rather than readings its documents.

[11] Sometimes trade secrets laws work -- the recipe for Coca Cola has remained a secret for decades. More often, trade secret law is ineffective. Trade secrets laws are hard to enforce in Silicon Valley and they are unenforced in much of the world. See Yuval Feldman, "Confidential Know-how Sharing and Trade-secrets Laws: Studying the Interaction Between Legality, Social Norms and Justice Among High-tech Employees in Silicon Valley" (PhD Thesis, Law School, University of California at Berkeley, 2004.

of positive profits launches innovative ventures, and the expectation of negative profits precludes innovative ventures. Figure 4 depicts this fact. The vertical axis represents venture profits.[12] The horizontal axis arrays ventures by profitability from high on the left to low on the right. With open competition, investors will finance ventures until profits fall to zero.[13]

Figure 4. Laws that Increase Venture Profits Cause More Innovations



Law affects the profitability of all phases of a venture -- finance, development, marketing, and competition. When better law makes innovations more profitable, the number of innovations increases. Figure 4 contrasts venture profits under two different legal regimes. Under the original law, venture profits reach 0 at innovation I. Improved law changes the situation by shifting the venture profit curve up as indicated by the arrows, so venture profits reach 0 at innovation I*. Thus improved law increases the number of innovations that get developed from I to I*. The figure depicts the increase in one period of time. If better law increases innovations in every future period of time, then the sustained growth rate increases.

---

[12] Venture profits equal the discounted present value of the stream of revenues from the innovation's sales, minus the discounted present value of its development and production costs

[13] Investors with private information will finance the highly profitable, infra-marginal ventures. Investors with public information finance the marginal ventures that earn zero profits. For marginal ventures, the present discounted value of future revenues equals the cost of development and production. The risk-adjusted return for investments based on public information is roughly the ordinary rate of return in technology industries and other industries.

II. Strength of IP

The patent for inventing a machine, creating a molecule, extracting a vaccine, or writing a computer program can be weak or strong. I will develop the theory and application of the optimal strength of patents. A patent's legal strength depends on three attributes: duration, breadth, and remedy. The *duration* of a patent can be measured in years. Thus U.S. patent duration is 20 years from the date of filing the patent application.[14] When filing, the applicant and her attorney make claims about the innovation's *breadth*. The patent office and the courts ultimately decide the validity of these claims. Breadth is easier to order than to measure. Thus a patent covering rain gear is broader than a patent covering umbrellas, and a patent covering umbrellas is broader than a patent covering automatically opening umbrellas. While one patent can be broader or narrower than another, there is no natural measure of the difference in breadth.[15]

Acting within the scope of a patent without a license from the owner infringes it. Law usually provides a remedy for infringement. For past infringements, the owner can usually sue to recover compensatory damages. Higher damages ("super-compensatory") are allowed sometimes, notably for willful infringement. For future infringement, the owner can usually enjoin the infringing conduct. The law can strengthen a patent by increasing its duration, breadth, or remedy.

One patent can strengthen by weakening another patent or weakening the rights of non-patentable goods. To illustrate, the space of innovations

---

[14] Before 1995, U.S. patent duration was 17 years from the date the patent issued.

[15] Patent breadth resembles the alphabet. Letters A through N encompasses more letters than A through M, but there is no natural measure of the distance between M and N. Similarly, no natural measure exists to answer the question, "Is the difference in breadth between a patent on rain gear and a patent on umbrellas larger or smaller the difference in breadth between a patent on umbrellas and a patent on automatically opening umbrellas?" In this respect, patent breadth contrasts with height. One person is taller than another and the difference can be measured in centimeters.

Fortunately, an ordering without a measure can support the mathematics of maximization. This fact makes modern utility theory possible in economics. Economic models often maximize utility functions that order states of the world by preference ("ordinal utility"), but do not measure differences in utility levels. Similarly, finding the breadth of patent that maximizes growth only requires an ordering by breadth, not a measure.

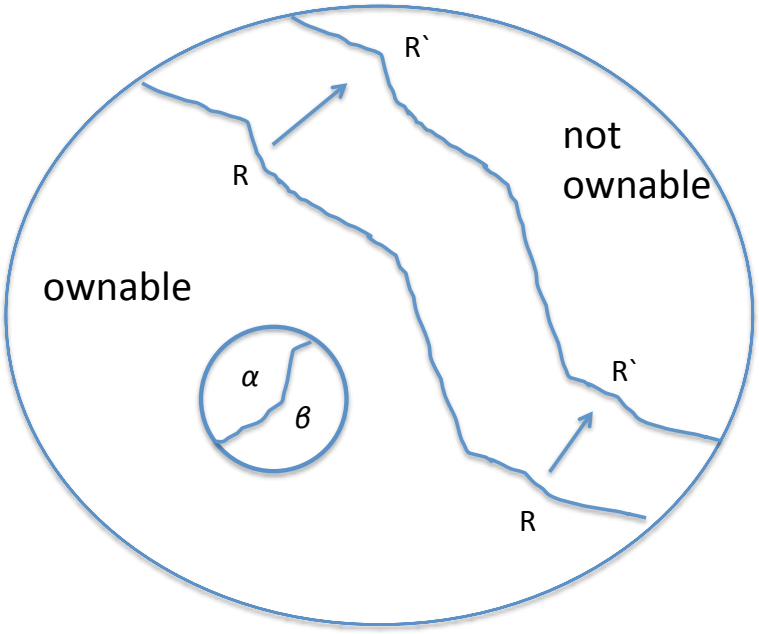
represented by the small circle in Figure 5 is subdivided into α and β. Assume that firm A invents α and firm B subsequently invents β. A claims that α and β are a single invention, and A files for a patent covering both of them. Later B files for a patent for β and claims that β is a different invention from α. The patent office must decide whether to award a broad patent to A that covers α and β, or to award a narrow patent to A that covers α and a narrow patent to B that covers β. For a given space of inventions, the state can generally give few broad patents or many narrow patents, rather like a city planner can divide undeveloped land into a few large lots or many small lots.[16]

In addition to strengthening one patent by weakening another, patent law can strengthen patents activities against unpatented activities. Figure 5 divides future innovations into "ownable" and "unownable."[17] A new computer chip is ownable (patentable) and a new metaphor in the English language is unownable (no patent, no copyright). The strength of all patents indicates the patent system's *reach*, as illustrated by the boundary between ownable and unownable innovations in Figure 5. Thus shifting the boundary in Figure 5 from R to `R increases ownable innovations and decreases unownable innovations. To illustrate, U.S. patent law increased its reach when inventors were allowed to patent business processes for the first time, such as Amazon's "one-click" ordering from its online catalogue. Expanding patent law's reach fences in more of the common land of innovation.

[16] Once lots are surveyed in a new town, owners can sell a lot but not half of a lot. The owner who cannot sell half of a lot can lease it for use by someone else. Much the same is true for selling patents. An owner cannot sell half of a patent, but an owner can contract for someone else to use half of it. Some theorists think that divisibility is the essential difference between property rights and contract rights. See Thomas A. Merrill & Henry Smith, "Optimal Standardization in the Law of Property: The Numerus Clausus Principle," 110 Yale Law Journal (2000).

[17] This is a picture, not a graph, because the breadth of innovations has no natural measure. See the discussion in this chapter of the difference between an ordering and an index.

Figure 5. Future Innovations



When a successful business venture develops an innovation, its launch yields high profits temporarily, then profits fall towards zero as the innovation disseminates and competition intensifies. Over its lifetime, the venture's profits equal a fraction of the innovation's social value – usually a small fraction – and the rest goes to other firms as profits and to consumers as "surplus."[18] Thus the wealth that Apple investors obtained from the iPhone is less than the value of the iPhone to consumers and other firms that imitate it or create applications for it. Since the social value of an innovation far exceeds the innovator's profits, venture profits provide deficient incentives to innovate.

[18] By definition, the consumer's surplus equals the difference between the price that a consumer would be willing to pay for a good and the price that he actually pays.

A. Separation Principle

What strength (breadth, duration, and remedy) of intellectual property rights is ideal? The answer is remarkably simple in principle. IP awards ownership of an innovation to its inventor. Owners of innovations in high demand with few substitutes have market power. Stronger property rights for innovators increases their market power. Innovators with market power can raise prices to buyers and enjoy extraordinary profits. By increasing the price of innovations, IP transfers wealth from buyers to sellers of innovations.

Innovations are bought for consuming, producing, and innovating. When market power transfers wealth from consuming and producing (static activities) to innovating (dynamic activity), the total profits of innovating increase. The marginal venture in Figure 4 shifts to the right, as depicted by the move from I to I*. By increasing the profits of innovating, IP increases the incentive to innovate, which causes faster growth. Thus market power for innovators against consuming and producing increase the rate of innovation. To promote progress in the useful arts, IP should give innovations strong property rights against consuming and producing.

When a firm uses market power to increase its price, part of the buyers' losses transfer to the seller as higher profits, and part of the buyers' losses do not transfer to the seller. "Deadweight" describes a loss without an offsetting gain.[19] To illustrate numerically, assume that a seller earns 0 profits under perfect competition and 100 under monopoly. A change from perfect competition to monopoly transfers 100 from buyers to the seller. According to standard economics, the seller gains less from monopoly than the buyers lose. Thus if monopoly causes the seller to gain 100, buyers lose, say, 130. The difference of 30 is the deadweight loss.[20]

---

[19] The deadweight loss occurs because the higher price causes buyers to purchase less of the good. When buyers forego some purchases, the buyers lose the gains that they would have enjoyed from these foregone goods, and the seller gains nothing from goods that buyers do not buy.

[20] The standard analysis is somewhat more complicated than these remarks suggest because of the difference between long run and short run effects. In the long run, all of the deadweight loss falls on

When patents create market power for innovations, the transfer of wealth from consuming and producing to innovating increases growth. The fact that prices increase for consuming and producing an innovation should not concern patent law for two reasons. First, the welfare gains from faster growth usually overtake the welfare losses from redistribution. Second, faster growth is the explicit purpose of patent law as given in the constitution.

Consider consumer prices of pharmaceutical drugs. Stronger monopoly powers for pharmaceutical innovators immediately increase the prices of drugs paid by consumers, which causes faster development of better drugs. Consumers lose immediately from costlier drugs and eventually gain even more from better drugs. Thus relatively cheap ulcer drugs replaced relatively expensive ulcer surgery. Similarly, cotton manufacturers buy ever-improving looms. Patents on looms increase the price of cotton textiles in the short run, but patents increase innovations that lower cotton prices in the long run.

I have been discussing the fact that strong property rights for innovations against consuming and producing increases growth. Besides producing and consuming, some innovations need prior innovations like a carriage needs wheels. When a prior innovation is owned, the owner usually has market power against a subsequent innovator who needs a license to use it. Market power of innovators against each other imposes a deadweight loss on all of them. When market power transfers wealth from one innovator to another, the total profits of innovators usually decrease, because they gain less in their role as sellers of innovations than they lose in their role as buyers. The difference between their gains as sellers and their losses as buyers equals the deadweight loss from transactions among innovators.

The deadweight loss decreases the total profits of innovators, so ventures devote fewer resources to innovation. The marginal venture in Figure 4 shifts to the left, as depicted by a move from I* to I. To promote progress in the useful

---

buyers, but in the short run sellers can also bear part of the deadweight loss. Note that in the standard graphical analysis, the transfer is a rectangle and the deadweight burden is a triangle.

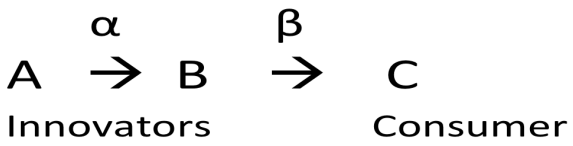
arts, IP should give innovators weak property rights, or no property rights, against each other.

Consider the licensing fee to use pharmaceutical drugs in research. Older drugs are often used to develop and test newer drugs. In a chain of such transactions, stronger IP increases the revenues from selling drugs to innovators by less than it increases the cost of buying drugs to make new innovations. Innovation slows, which harms consumers.

In sum, the ideal strength of IP depends on the activity burdened with higher prices. The *separation principle* asserts *the innovator's rights should be strong against others consuming the innovation or producing with it -- broad and long with a generous remedy for infringement*. Conversely, *the innovator's rights should be weak against others using the innovation to innovate – narrow and short with little remedy for infringement.*

Figure 6 depicts the separation principle. A creates innovation α and sells it to innovator B, who uses α to create innovation β and to sell β to consumer C. First consider the transaction between B and C. If B has market power and raises the price of β, then the dynamic economy as represented by B extracts resources from the static economy as represented by C, and the rate of innovation increases. According to the separation principle, law and policy should enhance innovator B's market power against consumer C.

Figure 6. Three-Party Straight-Line Transaction



Next consider the transaction between A and B. If A has market power against B, A can raise the price of α to B and capture some of B's profits from sales to C. Market power of A against B transfers resources from one innovator to another innovator. The deadweight loss from redistribution within the dynamic economy usually lowers its total profitability. According to the separation principle, law and policy should not usually enhance innovator A's market power against innovator B.

Innovators sometimes apply the separation principle when marketing innovations. Apple allows developers of applications to use the iPhone platform for free, whereas consumers must pay a fee for using many of the resulting applications. Thus Apple separates innovation from consuming, and asserts weak rights against innovating and strong rights against consuming. Similarly, Sun Microsystems developed the Java programming language in the 1990s and made it available for free to developers under a general-purpose license.[21] Developers use Java to write programs that they sell to consumers.[22] Thus Java is free for innovators and costly for consumers of its applications. Legal

[21] A reverse of this pattern sometimes occurs. Thus when the owner of a platform for running computer games allows consumers to use it for free and charges developers who want to use it to create new games to sell to consumers.

[22] Legal disputes subsequently arose when Microsoft modified java to run exclusively on its Windows program, thus necessitating the use of Windows to run programs written in the modified java language.

application of the separation principle would cause courts to mimic these licensing practices by private firms.

B. Fertility Principle

The separation principle has an important exception. An innovation is *fertile* like a breeding horse if it can be used to create another innovation. Today's new molecule is discovered from yesterday's new molecule; today's new operating system is discovered from yesterday's new operating system; today's new power cell is discovered from yesterday's new power cell.[23] Fertile innovations are often called "fundamental," "basic," "pioneering", "general," "first-generation," "rising technology," and "shoulders to stand on".[24] In contrast, an innovation is *sterile* like a mule if it cannot be used to create another innovation. Sterile innovations are often called "applied", "derivative," "specific," or "stand alone."

For purposes of economic analysis, one innovation is *more fertile* than another by definition if the costless transfer of profits from the latter to the former increases the growth rate. By definition, a transfer is costless if the loser loses the same amount as the winner wins, so deadweight loss is nil.[25]

If law and policy can transfer profits costlessly, then maximizing growth requires redistributing profits from less to more fertile innovations. Redistribution should proceed until all developed innovations are equally fertile. To illustrate by Figure 6, A's innovation α is used by B to make innovation β, so α is fertile.

---

[23] Brian Arthur describes technology as "self-creating" (autopoietic). "…if new technologies were constructed from existing ones, then considered collectively, technology created itself." W. Brian Arthur, The Nature of Technology (2009), page 2.

[24] See Timothy F. Bresnahan & Manuel Trajtenberg, 1995. "General Purpose Technologies "Engines of Growth?"," NBER Working Papers 4148, National Bureau of Economic Research; Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 JOURNAL OF ECONOMIC PERSPECTIVES 29–41 (1991).

[25] In a costless transfer from B to A, the burden of the transfer on B equals the benefit of the transfer to A. Mathematical notation makes this characteristic more precise. In notation, let i denote an innovation. The winner of an open competition to make innovation i enjoys profits denoted πi. The sustainable growth rate g is a function of the distribution of profits: $g=g(\pi 0,\pi 1,\pi 2,\ldots,\pi I)$. By definition, if j is "economically more fertile" than i, then a costless transfer of profits δ from i to j increases the growth rate:

$$g(\pi 0,\pi 1,\pi 2,\ldots,\pi I) < g(\pi 0,\pi 1,\pi 2,\pi i-\delta,...,\pi j+\delta,..\pi I).$$

However, B's innovation β is not used to make another innovation, so β is sterile. If IP gives A market power in the sale of α to B, wealth is transferred from the sterile innovation to the fertile innovation. If the transfer were costless, it would increase growth.

When innovators buy and sell innovations to each other, however, the transfer is usually costly, not costless. Market power imposes a deadweight loss on most transactions, including transactions between innovators. Deadweight loss among innovators causes growth to decrease. These considerations imply a general principle. The *fertility principle* asserts that maximizing innovation requires strengthening IP for more fertile innovations against less fertile innovations until the increase in growth from more fertility equals the decrease from more deadweight loss. (The mathematical expression is in a footnote.[26])

Creating market power for an innovation causes a large deadweight loss when the patented good has close substitutes (high elasticity of demand),[27] few obstacles to resale (no price discrimination),[28] the innovation is complex and opaque (incomplete information for buyers),[29] and many buyers purchase small

---

[26] In footnote 25, if j is a "economically more fertile" innovation than i, then a costless transfer of profits δ from i to j increases the growth rate:

$$g(\pi 0,\pi 1,\pi 2,\ldots,\pi I) < g(\pi 0,\pi 1,\pi 2,\pi i-\delta,\ldots,\pi j+\delta,..\pi I).$$

However, an actual transfer caused by law and policy creates a deadweight loss. The burden on i is δ, the transfer to j is γδ, and the deadweight loss is (1-γ)δ. At the optimum as given by the fertility principle, the benefit from the transfer to a more fertile innovation exactly offsets the cost from the deadweight loss, so the growth rate does not increase or decrease:

$$g(\pi 0,\pi 1,\pi 2,\ldots,\pi I) = g(\pi 0,\pi 1,\pi 2,\pi i-\delta,\ldots,\pi j+\gamma\delta,..\pi I).$$

[27] Goods are substitutes when having more of one makes a person demand less of the other. When a good has close substitutes, raising its price (while keeping other prices constant) causes buyers to switch to the substitute. Switching costs are a deadweight loss. Thus if a patent on computer-driven printers causes their price to rise, people will print less, which causes productivity to fall. Productivity falls more when the elasticity of demand for printers is high.When a seller uses market power to increase a good's price, buyers respond a lot if demand is elastic, and buyers respond a little if demand is inelastic. A larger response indicates that buyers are making more costly adjustments to cut back more on their purchases of the good.

[28] With few obstacles to resale, the owner cannot charge different prices to different buyers for the same innovation. Price discrimination is impossible. The more the seller can discriminate among buyers, the less the buyers will switch in response to a price increase, so the deadweight loss of switching falls. This is an application of the proposition found in microeconomics textbooks that a perfectly discriminating monopolist is efficient.

[29] As a good becomes more complicated and opaque, the amount of bargaining increases, and so do the search costs of determining whether an innovation is already patented.

amounts of the good.[30] If transactions among innovators satisfy these conditions, this fact argues for weak IP. If transactions among innovators do not satisfy these conditions, this fact argues for strong IP. The opposite of these conditions, and large differences in fertility, argues for strong arguments IP, as required to overcome the presumption that innovators should have weak intellectual property rights against each other.

C. Transaction Costs

Behind the deadweight loss are transaction costs that implicate deeper theoretical problems. In perfect competition, everyone takes the market price. In contrast, market power enables the seller to choose the price. The seller names the price and the buyers respond to it. Sellers with market power name prices in two different ways. First, the seller may name a *firm* price, as when the vendor in a football stadium sets the price of a hotdog. In the standard model of monopoly, the monopolist sets a firm price that maximizes profits.[31] Second, instead of a firm price, the seller may name a flexible price, and each buyer may make a counter-offer. With price flexibility, the parties bargain with each other to reach an exact price. Bargaining takes time and effort. In addition to time and effort, bargaining sometimes fails. Three major costs of transacting by bargaining are time, effort, and failure to agree.

Transaction costs are lost, not transferred. Consequently, "transaction cost", "deadweight loss", and "costly transfers" are almost equivalent terms. If there were no transaction costs, a monopolist could write contracts that eliminate the deadweight lost.[32] Reducing the deadweight loss requires charging different

---

[30] Transactions costs of purchases often have economies of scale. Consequently, the transaction cost of market exchange is higher with many buyers make small purchases than when few buyers make large purchases.

[31] At the profit maximizing price, the marginal revenue from a small increase in production equals the marginal cost. Thus a seller's revenue equals the price p multiplied by the quantity q that he sells. For a monopolist, the quantity is a decreasing function of the price. So the monopolist's revenue R can be written $R=pq(p)$ where $q`<0$. The monopolists cost of production C increase with the quantity, which decreases with the price, so $C=C(p)$ where $C`<0$. Profits, which equal revenues R- C, are maximized when marginal revenues equal the marginal cost: $q+pq`= C`$.

[32] Reducing deadweight loss of monopoly often requires replaces a single price with a price schedule that discriminates among buyers.

prices to different buyers. Thus the deadweight loss occurs because transaction costs make perfect price discrimination impossible.[33]

Congress asserted its constitutional power to make patent law by enacting Title 35 Sections 100-105 of the United States Code.[34] Regulations and court decisions increase the precision and certainty of these laws. Statutes and cases concern patentable subject matter, novelty, priority, obviousness, interference, equivalents, market abuse, inequity, and other topics. The separation and fertility principles can often clarify and critique these laws.

To apply the separation principle, lawmakers and courts must distinguish the uses of an innovation into consuming, producing, and innovating. In many cases, the distinction requires non-technical judgments of the kind that courts routinely make without statistics or mathematical theories. Hard distinctions occur at the boundaries of these three activities. Economist can help by predicting the effect of alternative interpretations on the rate of innovation. I will illustrate some applications of separability and fertility to specific IP laws.

D. Research Exemption

The common law historically permitted the unlicensed use of patented goods for non-commercial experiments and research by universities and other non-profit organizations. In common law, the experimental use defense followed the separation principle's prescription of treating research differently than consumption and production, with weak patents or no patents against researchers. Unfortunately, the courts sharply curtailed this defense.[35]

---

[33] By definition, a perfectly discriminating monopoly charges the reservation price to each buyer, so the marginal rate of substitution equals the marginal rate of transformation as required for static efficiency. With zero transaction costs, perfect discrimination eliminates the deadweight loss of monopoly. Thus price schedule with quantitative discounts discriminates according to how many units of the good the buyer buys, charging a higher price for the first unit and a lower price for the last unit.

[34] Section 101 [2] sets out subject matter that can be patented; section 102 [3] defines novelty and loss of right to patent; section 103 [4] lists what constitutes non-obvious subject matter.

[35] *Madey v. Duke University*, 307 F.3d 1351, 1362 (Fed. Cir. 2002), and *Roche Products, Inc. v. Bolar Pharmaceutical Co.* (Fed Cir. 2004).

Fortunately, Congress partly restored the research exemption for pharmaceuticals through the Hatch-Waxman Act (35 U.S.C. § 271(e)(1)). Firms that develop a new drug need regulatory approval before marketing it in the U.S. To gain regulatory approval, the drug's owner must prove that it is effective and safe. Proof often requires research on other chemically related drugs, including rivals for the same market. The owner of a drug has reason to refuse to license its use by a rival. Hatch-Waxman allows firms to use patented drugs for research on regulatory compliance without a license from the patent's owner.

The separation principle commends generalizing the research exemption beyond pharmaceuticals by restoring the historic common law rule.

E. Inventing versus Discovering

The boundary between owned and unowned innovations defines the intellectual property system's reach as depicted in Figure 5. Courts and regulators have attempted to draw the boundary in various ways. One line of cases distinguishes inventions that are patentable, and laws of nature that are not patentable.

To illustrate, the inventors of the telegraph in the 19th century attempted to patent the use of the electromagnetic spectrum to transmit messages. In *O'Reilly v. Morse*, the U.S. Supreme Court decided that the telegraph's inventors could patent a particular machine to transmit messages by using the electromagnetic spectrum, but not all ways of doing so. The electromagnetic spectrum is unpatentable, like the laws of nature and observations of physical phenomena.[36]

How does this legal distinction square with the separation and fertility principles? The separation principle favors strong patent protection over static activities (consuming and producing) and weak or no patent protection over dynamic activities (innovating). The telegraph was the first device to use the electromagnetic spectrum to transmit messages, and many others followed,

[36] The court recently sustained this view in *Bilski* (2010).

including telephones and computers. The patent for the telegraph gave its inventor's temporary market power over transmitting messages by using the electromagnetic spectrum, until alternative technologies created competition. Thus the patent was strong against telegraph transmissions, which mostly concerned consuming and producing. Conversely, the patent was weak against follow-on inventions like telephones. This result is broadly consistent with the separation principle, because it created profits for the original innovation and strong incentives for subsequent innovations

A similar conclusion concerns DNA sequencing. Does the first person to observe a DNA sequence acquire ownership of it, like Christopher Columbus claiming a continent when he discovered a Caribbean island? In *Myriad Genetics* (2013) the U.S. Supreme Court unanimously decided "No". In the same case, however, the Supreme Court allowed a patent for modifying DNA to produce a substance not found in nature. I apply the separation principle to this case. In order to innovate, many subsequent researchers will need to observe this genetic sequence. Thus denying a patent on observing DNA will lower the cost of future innovation.

In contrast, physicians will use the modified DNA to diagnose cancer in patients. Allowing a patent on modified DNA will reward the inventor and raise treatment costs to patients (consumption). Thus the Supreme Court's distinction in *Myriad* between unpatentable observation of DNA and patentable modification of it seems broadly consistent with the separation principle, although its explicit use would ground the decision better. This distinction between observing a DNA sequence and modifying it also applies to discovering natural laws and making inventions. The non-patentability of natural laws and the patentability of inventions seems broadly consistent with the separation principle, although its explicit use would ground the decision better.

*Prometheus (*2012) illustrates the problem with failing to use separation explicitly. The plaintiff performed exhaustive statistical analysis to develop benchmarks in standard blood tests for prescribing the best dosages for treating different patients. The plaintiff asserted that the use of its benchmarks in

prescribing drugs is a patentable method. The U.S. Supreme Court, however, disagreed. Unfortunately, Court did not consider the separation principle. Better benchmarks are innovations that benefit patients. According to the separation principle, innovation should have strong patents against consuming them, which implies that the creator of better benchmarks should be able to patent them.

The Court's decision in this case creates perverse incentives. Better benchmarks that improve the results of drug treatment are costly to create. Given the court's decision against patentability, doctors can profit most from improving benchmarks and keeping them secret. Instead of publicizing the benchmarks, the doctor can publicize the superior results obtained by using them, and charge higher fees to patients. In medical treatment, secrecy about improvement is socially destructive or even tragic. A court decision crafted around the separation principle would have allowed the patent and avoided the incentive for secrecy.

F. Novelty and Breadth

Besides being useful, U.S. patent law requires an invention to be new. An invention is new if it is unknown to the public and not obvious. Important disputes about novelty concern whether one innovation is sufficiently different from another to justify a separate patent. This is a question of the breath of patents as represented in Figure 5. Common sense and engineering principles give the answer in many cases, but not in the hard ones. Resolving hard cases requires connecting legal rules to the purposes of IP -- promoting economic growth through innovation.

Connecting patent breadth and growth requires understanding how economists measure similarity in goods. Economists regard two goods as similar if they are substitutes in markets, which implies that people will buy the one that is cheaper and not buy the dearer one (high cross-elasticity of demand). Recall Figure 5 in which A invents α and B subsequently invents β. If α and β are close substitutes for each other, and if no close substitutes exist for them, then one broad patent will create a monopoly and two narrow patents will create a

duopoly. Compared to duopoly, a monopoly can charge at least as high prices to buyers.[37] So one broad patent for A covering α and β in Figure 5 transfers more wealth from buyers to A than two narrow patents transfer to A and B.

When goods are substitutes, their use determines whether transferring wealth from buyers to sellers increases or decreases incentives for innovation. If innovations α and β mostly have *static uses* (consuming and producing), then one broad patent will transfer more wealth from static to dynamic activities than two narrow patents, which increases the incentive to innovate. Thus the separation principle favors broad patents for innovations that substitute in consumption or production. Conversely, if innovations α and β are substitutes and they mostly have *dynamic uses* (innovating), issuing two narrow patents will impose less deadweight loss on innovating, which increases the incentive to innovate. Thus the separation principle favors narrow patents for substitutes in innovation. (In the exceptional case of a very fertile innovation, however, the fertility principle favors a broad patent for substitutes in innovation.)

The conclusion is different for goods that are independent, rather than substitutes. Demand for two goods is independent if the price of one does not affect how much people will pay for the other (zero cross price elasticity). Independent goods are dissimilar economically. Intellectual property law cannot increase the market power or profitability of innovating by issuing broad patents encompassing independent inventions. Thus when demand for α is independent of demand for β in Figure 5, owning the patent on β does not increase the profitability of owning the patent on α. Patent authorities should issue narrow parents for inventions of independent goods because broadening these patents does not increase the profitability of innovating.

This reasoning applies to a series of suits that Apple brought against Samsung beginning in 2011 for infringement of design patents for the iPhone and iPad. For smart phones and tablets, Apple was the original innovator and

[37] Duopolists sometimes cooperate perfectly with each other, or they merge. In that case, they maximize their joint profits and earn the same profits as a monopolist. More often, however, duopolists cooperate imperfectly with each other and do not monopolize their joint profits. So the profits of a monopolist are at least as high as a duopolist for the same product.

Samsung was the imitator. Samsung's strategy was to build close substitutes to the originals and to make small improvements. By imitating, Samsung saved development costs, so Samsung built substitutes at lower cost and sold them to consumers at lower prices. Consequently, Samsung's share of the smartphone market increased at Apple's expense.[38]

The legal question is whether Samsung infringed Apple's patents. In so far as Samsung mostly used Apple's innovations to make consumer goods, the separation principle favors strong protection of Apple. Strong protection will reward innovation and speed progress in the useful arts. In so far as Apple's innovations were fundamental and Samsung's innovations were derivative, the fertility principle also favors strong protection of Apple's patents against Samsung. Again, strong protection will reward fertile innovation and speed progress in the useful arts.

However, in so far as Samsung mostly used Apple's innovation to make its own innovations of similar fertility, the separation principle favors weak protection of Apple's patents against Samsung. Weak protection will reduce the deadweight loss on innovation and speed progress in the useful arts.

Samsung's liability to Apple ideally depends on the extent to which Samsung's smart phones and tablets embody its own fertile innovations. Samsung's devices appear to use Apple's designs mostly to increase consumer sales, and Samsung's innovations appear to be derivative. If these appearances match the deep truth, then Apple's patent protection should be strong against Samsung. Liability of Samsung to Apple will promote faster growth in the future by causing Samsung to devote more of its formidable technological prowess to fundamental inventions and less to imitations.

As noted, IP awards ownership to the winner of the innovation race, and there is no second prize. However, the patent system can change the number of first prizes by redefining the breadth of patents. Decreasing the breadth of patents results in more patents for related inventions. More patents implies that

---

[38] Sang-Hun Choe, *After Verdict, Assessing the Samsung Strategy in South Korea By CHOE SANG-HUN* ``, NEW YORK TIMES, Summer, 2012, http://www.nytimes.com/2012/09/03/technology/companies/south-korea-reassesses-samsung-after-battle-with-apple.html?_r=1.

each patent holder has less protection against substitutes. In brief, decreasing the breadth of related patents causes more patents of less value. In aggregate, decreasing the breadth of patents moves away from monopoly and towards competition, so the aggregate profitability of patents decreases. Increasing the breadth of patents has the opposite effect. Applying the separation principle, innovation accelerates from increasing the breadth of patents against consuming and decreasing the breadth of patents against innovating.

Another legal problem of patent breadth concerns the independent invention defense. Suppose that two people independently make the same invention and only one of them receives the patent. The person without the patent infringes by using his own invention in his own laboratory. This often occurs in the course of secretive research because of perverse incentives in the law of damages.[39] The law must decide between holding the infringer liable or allowing an "independent invention defense", which excuses infringements by an independent inventor using his own invention in his laboratory. In effect, an independent invention defense narrows the breadth of the patent, which reduces profits. Applying the separation principle, an independent invention defense should distinguish between infringing to innovate and infringing to consume or produce. To promote growth, the independent invention defense understood broadly for infringing to innovate and narrowing for infringing to consume or produce.[40]

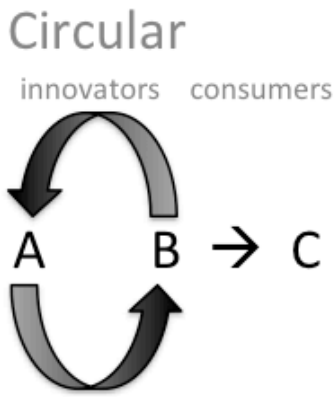
[39] The remedy for willful infringement is treble damages. Sometimes laboratory workers deliberately avoid informing themselves about patent owned by others because they prefer to infringe unknowingly and thus avoid treble damages.

[40] This is the beginning of an analysis of the independent invention defense. A complete analysis of the independent invention defense requires considering innovation races, which may cause competing ventures to duplicate research or to invest too much too soon. The problem of haste and duplication caused by the rule, "First in time, first in right", is discussed in chapter 5 of Cooter and Ulen, Law and Economics (6th edition). I will not rehearse the historical debate about whether economic planning can improve the allocation of investment by capital markets. For the famous Lange-Lerner debate, see LERNER, THE ECONOMICS OF CONTROL (1944), and OSKAR LANGE, ECONOMIC THEORY AND MARKET SOCIALISM: SELECTED ESSAYS OF OSKAR LANGE (1994).

G. Thickets

Much that a venture buys to develop an innovation was originally another venture's innovation. When innovators sell innovations to each other, some transactions circle within the dynamic economy, instead of proceeding straight from innovators to consumers as in Figure 6. In Figure 7 venture A makes innovation α and sells it to venture B, who uses it to develop innovation β and sells it to venture A for the next round of innovations, and so on. (In Figure 7, innovators A and B also sell innovations to consumers C.)

Figure 7. Circular Transactions



Thus assume that innovator A in Figure 7 uses market power to raise the price of the innovations that it sells to B. This use of market power gains profits of 100 for A and imposes costs of 130 on B. Similarly, assume that innovator B uses market power to raise the price of the innovations that it sells to A. This use of market power gains profits of 100 for B and imposes costs of 130 on A. The deadweight loss of monopoly in this cycle of transactions between A and B is 60. If market power were eliminated, the increase in profits of A and B would

total 60, with each one gaining 30. Their profits would increase because each gains more from lower costs than it loses from lower revenues.

Circular transactions among innovators are sometimes described as an innovation "thicket." In an innovation thicket, everyone needs the innovations of others in order to make their own innovations. One solution to this problem is illustrated by the computer operating system Unix that Bell Labs developed in the 1960s. Unix is "unowned" or "open access" or "in the public domain." Its users are unrestricted by patent or copyright, although it is trademarked.[41] Consequently, many different programmers created variants of Unix without needing to license it, including the version powering Apple computers (BSD developed at the University of California at Berkeley).

Since no one privately owns Unix, using it does not require negotiating with anyone. If someone owned Unix, however, anyone who wanted to improve it would have to license its use. Negotiating a license takes time and effort, which reduces profits and slows innovation. An innovation thicket contains so many circular transactions that bargaining over property rights slows growth. Innovations in a thicket ideally remain in the public domain where anyone can use them freely, like Unix. If not in the public domain, then innovations in a thicket should receive weak patent rights.

The more innovators deal in circular transactions with each other, the larger is the increase in growth from reducing their market power. Eliminating market power in circular transactions by private agreement is relatively easy when everyone gains. Three different kinds of private transactions among innovators can sometimes overcome the patent thicket problem. Large firms often accumulate a rich patent portfolio. When firm A needs to use a patent in B's portfolio, sometimes B needs to use a patent in A's portfolio, so they can negotiate *cross licenses*. Or, instead of waiting until the need arises, cross licensing can be negotiated in advance. Each member of a *patent pool* agrees in

[41] Only users conforming to certain specifications can apply the name "Unix".

advance to cross-license its patents for a given technology to everyone else.[42] Or, instead of forming a pool, owners of complimentary patents can *merge* into a single firm that holds all of the patents required to develop an innovation.[43]

Each of these solutions – cross licensing, pools, and mergers -- has limited success in solving the patent hold out problem.[44] In practice, firms cross license or pool a small fraction of patents, and small firms often resist merging for fear that size will impede creativity.[45]

Private transactions among individuals often fail in innovation thickets because of a bargaining problem. If innovations in a thicket are patented, then a venture to develop an innovation needs licenses from several other innovators. When several patent owners have the power to withhold essential licenses, then each of them must consent for the venture to proceed. As the number of patent-holders with veto power increases, the probability increases sharply that someone will block the venture. Similarly, many barons built castles on the Rhine River in medieval times to extract tolls from passing ships. One baron who held out for a high toll could block most traffic on the river. If law allows holdouts, they sharply reduce innovation in thickets and trade along rivers.[46]

A technical fact lies behind this problem. Assume that a venture can earn 500 if each of 5 individuals cooperate by supplying patents, but the venture can only earn 300 if 4 individuals cooperate and any one individual does not cooperate. The marginal product of a participant in a cooperative venture equals the increase in the venture's value when he participates in it, holding the participation of others constant.[47] Thus the marginal product of the 5th individual

[42] A subsequent chapter on contracts discusses cross licensing and pools.
[43] A subsequent chapter on antitrust discusses mergers among innovators.
[44] Each firm wants to give the pool its least valuable patents, and each firm wants to use the most valuable innovations of others in the pool. This is the problem of adverse selection.
[45] Merging dilutes the concentration of ownership needed to solve the double trust problem, as mentioned in Chapter 3 and examined further in a subsequent chapter on antitrust.
[46] Hold-out is a problem of last-mover advantage, which causes paralysis. The symmetrical opposite is the problem of first-mover advantage, which causes a stampede. To promote growth, law needs to cause a stampede of innovation, not paralysis.
[47] This is usually equivalent to the decrease in the venture's value when he stops participating in it (holding the participation of others constant).

to join the venture is 200. When distributing the value of a joint product, the usual rule in economics is for each participant to receive his marginal product. If everyone cooperates except for one, the hold out can credibly demand 200 to join the venture, leaving only 300 for the four others. Since each patent holder gains by consenting last, no one wants to consent first. Everyone holds out and hopes to be the last to join the venture, which fails to launch.

In general, a venture with increasing returns to scale cannot pay each participant its marginal product, because the sum of the marginal products exceeds the coalition's total product.[48] The venture does not have enough value for each contributor to receive his marginal product. If the participants negotiate and each one demands his marginal product, the venture is doomed.[49] Demands that are reasonable individually are unreasonable collectively. (Note that in these circumstances, marginal product also provides no basis for computing compensatory damages or reasonable license fees.[50])

Given its intractability, private actors need law's help to solve the hold out problem in patent thickets. Law can help by weakening patents among innovators in a thicket. Narrowing a patent's breadth weakens its market power and lowers the price that a license commands. In *Apple v. Motorola*, Judge Posner recently followed this prescription by dismissing Apple's infringement claims. Law can also help to solve the holdout problem by weakening the remedy for patent infringement. Thus in *eBay v. Merc*, the U.S. Supreme Court held that the weaker remedy – damages instead of injunctions – may apply to small components in a complex innovation.

---

[48] In the language of game theory, a coalition game with increasing returns to the size of a coalitions has an empty core.

[49] In more technical language, increasing returns to the scale of a coalition implies that, and the game's "core is empty".

[50] Damages for using another's property often equal the marginal product that the owner lost when the infringer used his property (compensatory damages), or else damages equal the marginal product that the trespasser got from using another's property (disgorgement damages). In a thicket, "compensatory damages" or a "reasonable fee" for using a patented good without a license cannot refer to its marginal product, because the sum of the marginal products exceeds the total product. Hence marginal productivity theory provides no way to compute damages from infringement of patents in a thicket, nor for negotiating the shares from cross-licenses.

H. Middlemen and Trolls

"Middlemen" refers to intermediaries who buy goods for resale. Throughout history, people have viewed middlemen with suspicion because they make money without making goods. Confusion abounds about their economic role. Do they increase or decrease prices to consumers? Do they create cartels or match buyers and sellers?[51]

The same confusion infects discussions about patent middlemen, who buy patents to enforce them, not to produce or innovate. According to the Norwegian tale of the "Three Billy Goats Gruff", a troll living under a bridge gobbled up pedestrians who tried to cross. Similarly, some law firms gobble up patents and sue others who need them for production or innovation. "Patent troll" is the derogatory term for "patent middlemen."

Name calling aside, law requires analysis of patent middlemen. After a business venture develops and patents an innovation, the innovation must be marketed to users and enforced against infringers. With respect to enforcement, the innovator can directly defend her patent or she can indirectly defend by selling the right to collect damages to someone else. Thus firm A invents and patents α. A can enforce the patent on α, or A can sell α to B, or A can sell to B only the right to sue and collect damages from infringers of α.

Innovating, marketing, and enforcing are different specialties. Different people and firms have a comparative advantage in different activities. Some middlemen like Intellectual Ventures or MPEG LA aggregate large portfolios of intellectual property, and a user who needs several licenses can buy them all at once from the aggregator.[52] The aggregator reduces the cost of acquiring

[51] Middlemen often save buyers and sellers the cost of finding each other, but they sometimes form cartels to separate buyers and sellers, and to extract higher prices. Thus middlemen try to prevent buyers from obtaining information about sellers from any source but themselves. To illustrate, the American Bar Association historically prohibited its members from advertising their specialties, including printing a specialty on a business card. Consequently, a person who needed, say, a patent attorney often had to ask the family attorney for a reference. This restriction on information generated referral fees for lawyers. Such predatory practices by middlemen cause hostility and resentment among those who must pay for it. The U.S. courts eventually struck down such legal rules as violations of "commercial free speech".

[52] Thanks to Professor Ryan Holte for explaining the operation of these companies to me.

intellectual property and enforcing intellectual property rights. By facilitating middlemen, the law gives patent owners more distribution and enforcement choices. Innovators tend to choose the most profitable form of enforcing patents, whether doing it themselves or hiring others to do it for them. By lowering the costs of enforcement, more enforcement choices strengthen the patent owner's rights *de facto*, if not *de jure*.

Stronger patent rights can stimulate or stifle innovation, as previous discussions show. According to the separation principle, increasing a patent's strength against consumers and producers stimulates innovation. By prosecuting infringers who use an invention to produce consumer goods, patent middlemen transfer resources from consuming to innovating, which speeds progress in the useful arts. To speed progress in the useful arts, law should facilitate sales of patent rights to middlemen for innovations used in consumption or production.

Conversely, by prosecuting infringers who use an invention to innovate, patent middlemen impose a deadweight loss on innovators, which slows progress in the useful arts. To reduce this drag on innovation, law should usually inhibit sales of patent rights to middlemen for innovations used in innovation, as in innovation thickets. (Note, however, that the unusual case of a very fertile invention is an exception. For fertile innovations, law should facilitate sales of patent rights to middleman, according to the fertility principle.)

I. Duration

Next, apply the separation and fertility principles to the duration of a patent. The horizontal axis in Figure 8 indicates the length of patents in years, and the vertical axis indicates the rate of innovation. The curves in Figure 8 depict some possible relationships between the rate of innovation and patent length. For curve A, the rate of innovation is an increasing function of patent length. For a curve like A, innovation-maximizing patent length is infinite. Curve A might depict an industry where innovators sell mostly to consumers and

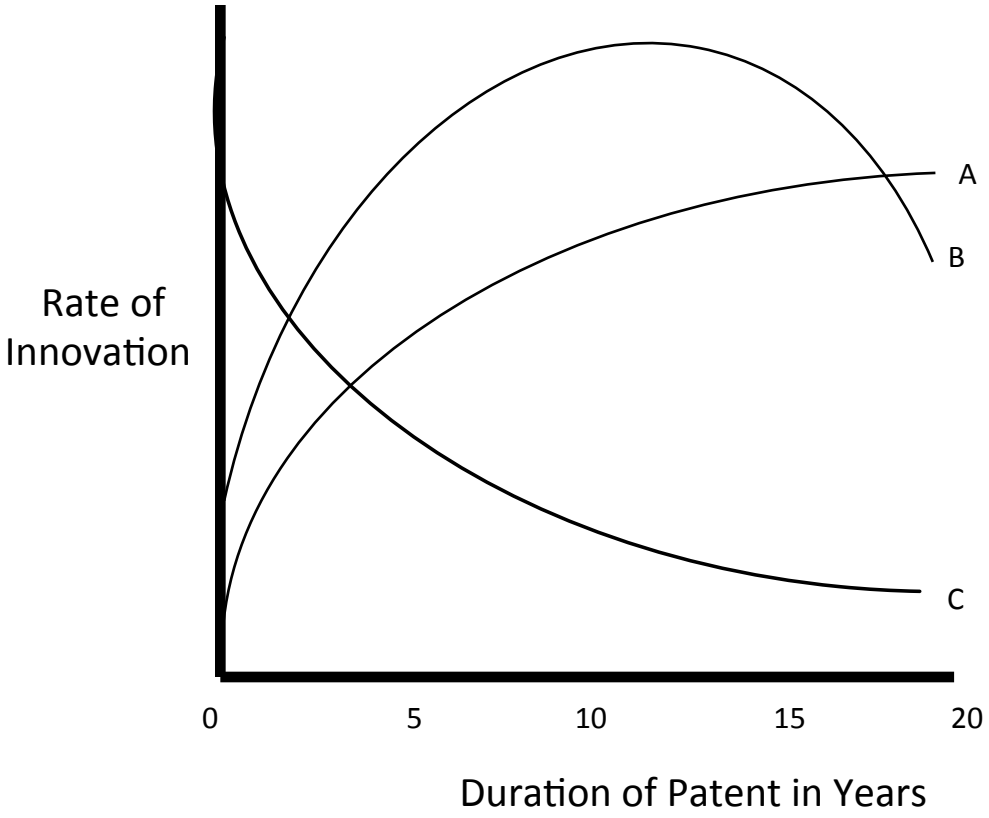
producers, such as pharmaceutical drugs where the optimal patent is very strong.[53]

For curve C, the rate of innovation is a decreasing function of patent length. For a curve like C, the innovation-maximizing patent length is zero. Curve C might depict an industry where innovators sell mostly to each other, and their fertility is similar, so the optimal patent is very weak. In these circumstances, the dominant effect of longer patents is to transfer more wealth among dynamic users with deadweight loss.

Curve B depicts the intermediate case in between the pure cases of Curve A and Curve C. Starting with weak IP rights to the left on Curve B, moving to the right and strengthening IP increases the rate of innovation at first. A point is reached, however, where further strengthening decreases innovation. In Figure 8, the innovation-maximizing patent length for Curve B is internal, neither infinite as for curve A nor zero as for curve C. In the intermediate case, moderately strong intellectual property rights maximize the rate of innovation.

Figure 8. Rate of Innovation and Patent Length

[53] Landes and Posner advocate patents of infinite length, which we favor for patents against consuming and producing, but not for patents against innovating. CITE;



Curves A, B, and C in Figure 8 depict three possible relationships that often exist between the rate of innovation and the duration of intellectual property rights. (I will not discuss other possible shapes.[54]) For curves with these shapes, a simple rule leads to the maximum: *Starting from 0, increase the strength of intellectual property rights so long as venture profits increase.*

The ideal patent system grants patents of different length according to the growth profile as in Figure 8. German patent authorities issue "petty" patents of short duration for minor inventions, and "full" patents of longer duration for major inventions. In contrast, American patent law issues full patents for *all* patented

[54] Curve B is a concave function with a unique maximum. Another possibility is a convex function with two local maxima, one for weak intellectual property rights and one for strong intellectual property rights. Another possibility is multiple equilibria. For analyzing many activities, economists assume a concave function with a unique maximum.

inventions. Does the German or American patent regime come closer to this ideal? For inventions mostly sold for producing and consuming, full patents will transfer more wealth from static to the dynamic uses. Consequently, American patent law is closer to the ideal than German patent law for innovations mostly used for consuming and producing.

In contrast, the ideal law issues weak patents or no patents in the usual case for innovations used to innovate. In the exceptional case of very fertile innovations, the ideal law issues strong patents for innovations used to innovate. In so far as German patent authorities issue "full" patents to fertile inventions and "petty" patents to sterile inventions, German patent law is closer to the ideal than American law for innovations mostly used for innovating.

The international default standard for patent duration, which the U.S. has adopted, is twenty years from the date of the application for a patent. Quantitative research on the optimal duration of intellectual property rights has barely begun. No compelling evidence indicates whether the pace of innovation would increase from lengthening or shortening the duration, or from distinguishing between petty patents and full patents. (In contrast, copyright law probably retards progress in many of the useful arts because a copyright endures for so long – creator's life plus 70 years.[55] A specific U.S. statute on semiconductor protection possibly has a better rule.[56])

J. Remedy

Patent law provides damages for past infringements. Three benchmarks for damages are compensation, disgorgement, and treble damages.

---

[55] In October, 1998, Congress passed the Sonny Bono Copyright Term Extension Act, which lengthens copy- right protection for works created on or after January 1, 1978, to the life of the author plus 70 years, and extends existing copyrights "created for hire and owned by corporations" to 95 years. Before the change, the 1976 Copyright Act had given protection for the author's life plus 50 years. Whatever other reasons there may be for the Copyright Term Extension Act, one justification is that it brings U.S. practice into conformity with Western European practice. Many scholars believe that this is far too long.

56 The Semiconductor Chip Protection Act of 1984 (or SCPA) legally protects the layouts of integrated circuits for 10 years from registration. This act covers all embodiments of the chip's layout, but does not cover the chip's functionality that patents law protects or the information on chips that copyright law protects (if they are protected at all).

Compensation ideally puts the *victim* in the same position as if the infringement had not occurred. For firms, this requires transferring profits from the injurer to the victim until the *victim's* profits equal what they would have been but for the infringement. In U.S. law, compensation is the usual remedy for past infringement of intellectual property rights. (Compensatory damages differ in strength according to their basis of computation, as explained in a footnote.[57])

In contrast, *disgorgement* ideally puts the *injurer* in the same position as if the infringement had not occurred. For firms this requires transferring profits from the injurer to the victim until the *injurer's* profits equal what they would have been but for the infringement. In U.S. law, disgorgement is seldom available except when infringement of an intellectual property right involves breach of a fiduciary duty or unjust enrichment.

In addition to damages for past infringements, the owner can often obtain an injunction against future infringements.[58] If the law allows the owner to defend his patent with an effective injunction, then others must either desist from using it or pay the owner's *asking fee* for a license. Alternatively, instead of enjoining before infringement, the owner can collect damages after infringement. The owner with the right to an injunction can choose between enjoining infringer before the fact or obtaining damages after the fact. Having this choice of remedies increases the power of the patent owner relative to only having the damage remedy.

---

[57] Compensation ideally equals the decrease in the patent owner's profits caused by the infringement. The owner's profits equal the difference between revenues from selling the innovation and the cost of producing and developing it. Sometimes the cost of the invention's production and development is easier to document and prove than revenues foregone by sales that did not occur because of the infringement.

[58] In addition to patent law, U.S. firms have an additional remedy against infringing imports. The U.S. firm can bring a case before the U.S. International Trade Commission asking it to ban the import of all infringing goods. This powerful remedy against foreign infringement allegedly leads to abuse by firms shielding themselves from foreign competition. See K. William Watson, *Still a Protectionist Trade Remedy: The Case for Repealing Section 337*, No. 708 POLICY ANALYSIS (CATO INSTITUTE) (2012).

Treble damages simply means three time compensatory damages. Treble damages are the remedy in U.S. patent law for *willful* infringement of intellectual property rights.

The following inequalities order different remedies for infringement from weak to strong:

compensation $\leq$ compensation or injunction $\leq$ treble damages.

To maximize growth, how strong should the remedy be? The separation principle distinguishes between static (consumption and production) and dynamic (innovation) activities. Infringements for static uses should have strong remedies, in order to transfer more resources from static to dynamic activities and increase the pace of innovation.

U.S. law does not *always* give the owner the alternative of enjoining future infringements. Sometimes the law allows damages but not injunction. With damages, the user may infringe and subsequently pay a price set by the court. The price set by the court does not necessarily equal the owner's asking price.[59] Thus in *eBay v. Merc*, the U.S. Supreme Court held that when a patented invention is a small component in a large product, compensatory damages can compensate sufficiently for future use of the patented good, and no injunction will be given by the courts.

As explained above, a holdout problem occurs when making an innovation requires several patented components with different owners. Each owner cannot get paid his component's marginal contribution to the innovation.[60] By denying the right to an injunction, the Court in *eBay* effectively compels the owner of a patent on a small component in a large product to license its use at a price close to compensatory damages. The absence of an injunctive remedy prevents each

[59] Legal scholars say that property rights cannot be taken without paying the owner's *subjective price*, whereas the law allows other rights to be taken by paying an *objective price*. An implication of this fact is that sellers will not sell property for less than it is worth to them, and buyers will not buy property for more than it is worth them. Consequently, voluntary sales move property from people who value it less to people who value it more, which yields a surplus that can benefit buyer and seller. By creating surpluses, sales allocate resources to their most productive uses and reward innovators.

[60] Recall that with increasing returns from the use of components, the sum of the marginal contributions of the components exceeds the value of the innovation that needs them

owner of a component from demanding a licensing fee equal to its marginal contribution. Thus replacing the injunctive remedy with court-set damages for future infringement can avoid the holdout problem.

Unfortunately, the Court has not explained a general principle for setting damages in these circumstances.[61] According to this article, the ideal rule would set damages at a level that maximizes the rate of innovation. This approach would respond to the separation and fertility principles. Thus infringing to innovate would trigger a weaker remedy than infringing to consume or produce, and infringing a more fertile patent would trigger a stronger remedy than infringing a less fertile patent.

Conclusion

Intellectual property law creates an open competition to innovate. The innovator gets the first prize -- ownership of the innovation. There is no second prize. A property right conveys market power to owners of innovations in high demand with few substitutes. Market power creates extraordinary profits that economist's call "rents." Thus the law determines the ownership rights of innovators and the market determines their resulting rent.

Increasing the incentive to innovate speeds progress in the useful arts. Innovators enjoy a small fraction of the value that they create. So increasing the incentive to innovate increases the general welfare. Stronger property rights for innovators can increase or decrease the incentive to innovate. Innovations are bought for consuming, producing, and innovating. When market power transfers wealth from consuming and producing (static activities) to innovating (dynamic activity), the total profits of innovators increase, which increases the incentive to innovate. To promote progress in the useful arts, patent protection should be strong against infringing by consuming or producing. Unlike rents in politics and regulation, rents paid by consumers and producers to innovators increase wealth and welfare.

[61] The ideal rule, according to this article, would set damages at a level that maximizes the rate of innovation, which depends on the separation and fertility principles.

Conversely, when market power transfers wealth from one innovator to another, the total profits of innovators usually decrease, because they gain less in their role as sellers of innovations than they lose in their role as buyers. Like rents in politics and regulation, rents paid by innovators to each other decrease wealth and welfare. Consequently, patent protection should usually be weak or absent against users who innovate with the patented product. The *separation principle* asserts that maximizing innovation requires strong IP against consuming and producing, and weak IP against innovating.

This prescription has an important qualification. "Fertility" refers to an innovation's power to cause other innovations. Transferring wealth from less fertile innovations to more fertile innovations can increase growth. Patents should be strong against other innovators where the patented invention is more fertile than the innovation for which it is used. The *fertility principle* asserts that maximizing innovation requires strengthening IP for more fertile innovations against less fertile innovations until the increase in growth from more fertility equals the decrease from more deadweight loss.

The separation and fertility principles show how to implement IP's constitutional purpose through laws made by Congress and courts.[62] By raising the prices of innovations to consumers and producers, IP fulfills its constitutional purpose of progress in the useful arts. Conversely, by raising the price of innovations to innovators, IP frustrates its constitutional purposes, except when the winners are much more fertile innovations than the losers.

The separation and fertility principles can explicate and critique current issues in patent law, such as the difference between inventing and discovering, novelty and breadth, middlemen and trolls, and damages and injunctions. (Copyright is a related subject for consideration on another occasion.[63]) These

[62] Melvin Eisenberg explains the legal role of intermediate principles in The Nature of the Common Law (1988).

[63] The main difference in the economic foundations of patents and copyright concerns exponential growth. The welfare overtaking principle applies to innovations that can grow exponentially. The subject of patents is inventions, where progress often compounds. The subject of copyright is expressions, where progress sometimes compounds. Some copyrighted expressions such as computer programs can grow exponentially, as well as some innovations covered by special statutes. (The Semiconductor Chip Protection Act of 1984,

two principles simplify the interpretation of patent law because they do not require balancing growth against static losses or income redistribution. Balancing is not required by IP's specific constitutional purpose (innovation), or for IP's general constitutional purpose (national welfare), because the welfare gains from exponential growth overtake the static losses from inefficiency or redistribution.

---

which legally protects the layouts of integrated circuits, exemplifies such special statues.) In contrast, any attempt to measure progress in artistic expression (poems, novels, songs, plays) immediately encounters controversy. Artistic expression evolves and develops, but people disagree over whether development constitutes progress or the rate at which progress occurs. Consequently, the overtaking principle does not apply readily to these areas of copyright. Thus the economic interpretation of "progress in the useful arts" as maximizing economic growth applies to copyright more than patents.