

Intellectual Property: The Law and Economics Approach

Richard A. Posner

The traditional focus of economic analysis of intellectual property has been on reconciling incentives for producing such property with concerns about restricting access to it by granting exclusive rights in intellectual goods—that is, by “propertyizing” them—thus enabling the owner to charge a price for access that exceeds marginal cost. For example, patentability provides an additional incentive to produce inventions, but requiring that the information in patents be published and that patents expire after a certain time limit the ability of the patentee to restrict access to the invention—and so a balance is struck. Is it an optimal balance? This question, and the broader issue of trading off incentive and access considerations, has proved intractable at the level of abstract analysis.

With the rise of the law and economics movement, the focus of economic analysis of intellectual property has begun to shift to more concrete and manageable issues concerning the structure and texture of the complicated pattern of common law and statutory doctrines, legal institutions and business practices relating to intellectual property. Among the issues discussed in this paper are the length of protection for intellectual property, the rules that allow considerable copying of intellectual property without permission of the originator, the rules governing derivative works, and alternative methods of providing incentives for the creation of intellectual property. The emphasis is on copyright law, which, perhaps because of its complex legal structure and the relative neglect by economists of the arts and entertainment, has tended to be slighted in the conventional economic analysis of intellectual property, relative to patent law, where economic analysis can draw on an extensive literature concerning the economics of innovation. I also

■ *Richard A. Posner is a Judge of the U.S. Court of Appeals for the Seventh Circuit and a Senior Lecturer, University of Chicago Law School, both in Chicago, Illinois.*

touch on trademarks and trade secrets, and explore some parallels and contrasts between the legal treatment of physical and intellectual property. The subject is too large to admit of comprehensive treatment, however; for additional discussion and many references, see Landes and Posner (2003).

The Optimal Term

The tension between incentives and access that preoccupies the conventional economic analysis of intellectual property arises from the high ratio of fixed to variable costs of such property. Intellectual property is often very costly to create, but the costs of creation, being invariant to output, are fixed costs once incurred. In contrast, the costs that vary with output, which is to say the costs incurred in actually providing the intellectual property to consumers, often are very low, at least relative to the fixed costs; in the case of software distributed over the Internet (including digitized musical recordings), variable cost, and hence marginal cost, are close to zero. When fixed costs are a high percentage of total costs, a price equal to marginal cost is unlikely to cover total costs unless marginal cost is sharply rising. But a price above marginal cost, though necessary to enable the producer of the intellectual property to recoup fixed costs (unless those costs are subsidized), not only will deflect some potential purchasers to substitutes that cost society more to produce on a quality-adjusted basis, but it will also induce inefficient entry by firms that do not have to incur heavy fixed costs, as is commonly the case when a new entrant can free ride on the investment made by incumbent firms.¹ Marginal-cost pricing would maximize access to existing intellectual property and deter or expel inefficient entrants, but it would reduce, indeed often eliminate, the incentive to create the property in the first place.²

The principal alternatives for resolving the tension are a system of financial rewards to creators of intellectual property (such as a public subsidy) and a limited property-rights system (like patents and copyrights) that enables the creator of intellectual property to exclude others from access to it without the creator's authorization (Shavell and van Ypersele, 2001), but not to exclude as completely as in the case of physical property. A reward system, in principle, provides both incentives and access—the creator of intellectual property is compensated for the cost of creation, but because the creator has no right to exclude others from access

¹ A vital qualification: if the new entrant would have to incur heavy fixed costs, too, the entrant's average total costs might be no lower than those of the incumbent firm, in which event entry would not occur.

² A further complication, pointed out in a classic article by Kenneth Arrow (1962), is that because the market value of intellectual property often cannot be estimated in advance, risk aversion may result in the underproduction of such property from a social standpoint. It is not clear whether this problem is important, however. Risk can be reduced or even eliminated, often at low cost, by diversification, as where a book publisher obtains a normal return by publishing many books, some of which are flops but others of which exceed expectations.

to the property, the price is forced down to marginal cost by competition from copiers. Calculating the optimal reward is difficult, however. It can in principle be calculated after the fact on the basis of the commercial success of the intellectual property, but the resulting delay in reimbursement of the originator will increase the difficulty of financing intellectual-property projects.

A greater danger is that the reward system will be politicized. The danger is avoided when the system is private rather than public; the salaries that private universities pay professors who devote a substantial part of their working time to producing unremunerative intellectual property is an example of a private rewards system, as are Pulitzer prizes for writers.

The property rights approach proportions the creator's return on investment to the commercial success of the invention (in the case of patents) or expressive work (in the case of copyrights) automatically. But it may generate a return that exceeds the costs of creation even after adjustment for the risk of failure—the frequent “dry holes” in inventive and other creative activity—and may therefore restrict access unnecessarily. By the same token, as Arnold Plant (1934) long ago noted in classic articles that discussed “rent seeking” before the term existed, by enabling pricing in excess of marginal cost, intellectual property rights attract resources into the creation of such property that may have a larger social product in alternative uses in which, however, they would generate only a competitive return for producers.

Unfortunately, economists do not know whether the existing system of intellectual property rights is, or for that matter whether any other system of intellectual property rights would be, a source of net social utility, given the costs of the system and the existence of alternative sources of incentives to create such property.

Whereas the “pure” economist, whose approach I have been sketching, is likely to approach the question of optimal regulation of intellectual property from the standpoint of the economics of innovation, public goods and marginal-cost pricing, the economic analyst of law is more likely to begin with the parallels between intellectual and tangible property.³ The contrast in approaches can be illustrated with reference to the question of the optimal length of the copyright term. Once it is determined that the incentive benefits of propertizing expressive works exceed the access costs, the length of the copyright term—the duration of the property right—becomes a device for preventing the copyright owner from being overrewarded. This can be done by fixing the term at the point at which the expected ratio of the copyright works's total revenues to total costs (including opportunity costs), both discounted to present value, is 1. This formula for the optimal term usefully implies that if the term is already long, further increases are unlikely to affect the ratio, because of discounting to present value. The Sonny Bono Copyright Term Extension Act of 1998 increased the length of the copyright term from

³ For a summary of the economics of physical property, see Posner (2003, pp. 31–91).

the life of the creator plus 50 years to the creator's life plus 70 years. Such a change would be expected to have only a trivial effect on the present value of the expected revenues of creators of expressive works.⁴ It thus would have only a trivially positive incentive effect.

Less obviously, neither would the extension have a significant negative effect on access to the intellectual property, because future costs as well as future revenues must be discounted. But this conclusion is altered when, as has become customary for public-choice reasons considered later (and was true for example of the Sonny Bono Act), an increase in the copyright term is made retroactive—that is, applicable to existing copyrighted works as well as to new ones. The incentive effect with regard to previously created works is of course not only slight but nil. But access costs rise from the date on which the copyrights on existing works would have expired had it not been for the extension; and on some of those works the copyright would have expired the day after the effective date of the extension. Thus, attention to the legal details—in this example, to whether a law is made retroactive rather than merely prospective—can illuminate the economics of intellectual property.

Comparison with the law of physical property provides richer explanations than conventional economic analysis for why some types of intellectual property right, namely patents and copyrights, are time-limited rather than, as in the case of physical property, perpetual, while other types of intellectual property right, such as trademarks and trade secrets, do not have any definite term. Of particular importance in the case of both inventions (which are patentable) and expressive works (which are copyrightable) is the fact that the intellectual public domain, consisting of inventions that are not patented and expressive works that are not copyrighted, is a source of vital inputs into the creation of subsequent intellectual property. The temporal limitation of patents and copyrights augments the public domain because once the patent or copyright expires, the patented or copyrighted work cannot be reappropriated; it is forever a part of the public domain. The incremental character of the inventive process is obvious; but it is also the case that most creators of expressive works—whether novels, films, musical compositions, paintings or works of nonfiction, such as histories—borrow very heavily from earlier expressive works. In the words of the literary critic Northrop Frye (1957), “[P]oetry can only be made out of other poems; novels out of other novels.” Thus,

⁴ Suppose a copyrighted work was expected to yield revenue of \$1 per year in perpetuity at a discount rate of 10 percent. Under a system of perpetual copyright, the present value of this infinite stream of income would be \$10 ($= 1/r$). Under a limited copyright term ($= t$), the present value would be $(1 - e^{-rt})/r$. So even if $t = 25$, which is much shorter than the current term, the present value of the copyright (at $r = .10$) would be \$9.08, which is more than 90 percent of the present value of the perpetual copyright. For a work by an author who died at the age of 80 and had written the work at 40, the copyright term is 90 years under a rule of life plus 50 years ($80 - 40 + 50$) and 110 years under a rule of life plus 70 years. The present value of a 90-year copyright that yields \$1 a year in revenue is \$9.998, rising trivially to \$9.9997 for the 110-year copyright.

even creators of intellectual property might well accept a temporal limitation on their intellectual property rights, since they would benefit from having a larger public domain on which to draw without having to negotiate permission to copy.

Creators of physical property would not accept such a limitation. There would be no efficiency gains from a rule that after 25 years your house enters the public domain and anyone can take its walls and roof for use in building their own home. The difference lies in the much higher costs of transacting in intellectual property. Because the system for recording copyrights is primitive compared to that for recording titles to land and other tangible goods, the longer a work remains under copyright the greater is the cost of locating the copyright owner—and so the greater are the transaction costs of obtaining permission to copy the work. A longer copyright term will thus raise the average as well as the total transaction costs involved in the creation of new expressive works. In the case of patents, the longer the patent term, the more likely the invention space is to be cluttered with patents, requiring multiple negotiations and creating potential holdout problems. The public domain provides inputs that can be obtained without any transaction costs.

Yet long copyright terms also have economic benefits. (The benefits of long patent terms are less clear, and so one is not surprised that patent terms are considerably shorter.) A first such benefit is to prevent congestion, a common consequence of making access to property free to all users; a standard example is traffic congestion on nontoll roads. (So here is another example of fruitful application of the economics of tangible property to intellectual property.) Were “Mickey Mouse” in the public domain, the resulting surfeit of copies might produce a net reduction in the market value of the character if overexposure induced a degree of boredom or even disgust that caused, via a downward shift in the demand curve, a decline in total utility. Second, the dichotomy between the cost of creating an expressive work, which is incurred up front, and the cost of disseminating it, incurred after it is created, can be overstated. Maintaining a market for “Mickey Mouse” requires continued expenditures on advertising and promotion and continual minor modifications in the character itself designed to maintain the character’s appeal in the face of changing tastes. These expenditures may not be forthcoming if the character is freely copiable.⁵

The net effect of these additional tradeoffs on the optimal copyright term is an unresolved—indeed, unaddressed—empirical issue. A further complication is the arbitrariness of assuming that the same term for *all* expressive works is optimal. Given the heterogeneity of such works, a fixed term is likely to overreward some producers and underreward others, causing a misallocation of resources, although the copyright term is now so long that it is doubtful that any producer is being underrewarded by virtue of the durational limit on copyrights.

I mentioned that trademarks and trade secrets do not have fixed terms. There

⁵ Some of these modifications, however, may be copyrightable as derivative works, of which more later.

are good economic reasons for this. A trademark is a method of economizing on consumer search costs by providing a compact identifier of a particular producer's brand. So long as the producer continues to sell the brand, maintaining a more or less consistent quality so that consumers are not deceived, the trademark serves its identifying function. Competitors are not harmed, because a producer is not allowed to pick as its identifier a term or symbol that the public associates with the entire product (a generic term, in other words), rather than a specific producer's brand; for example, Microsoft could not use "Operating System" as the trademark for its operating system, because that would raise the costs to competitors of communicating with their customers.

Trade secrets have no fixed term, because the only legal protection that trade secrecy confers is protection against unlawful appropriation of confidential information, as by theft or a breach of contract. A competitor is free to appropriate a trade secret by reverse engineering or independent discovery, or because the firm that possessed the trade secret failed through negligence or inadvertence to keep it secret.

Transactions Costs, Fair Use and Piracy

The optimal term for protecting intellectual property may have received disproportionate attention from economists. By looking more closely at the legal rules governing intellectual property, law and economics analysis may be able to point the way to legal reforms that would be desirable regardless of the optimal term.

One reform, responding to the problem of transaction costs discussed in the preceding section, would be to require copyright owners who wanted to enforce old copyrights (say, copyrights issued a half century or more ago) to reregister them in a form that would make it easy for creators of new expressive works to identify the current copyright owner. Failure to reregister would be treated as abandonment, and the work would go into the public domain and could be freely copied. The cost to copyright owners would be trivial; very few copyrights old enough to be affected by the suggested reform have any commercial value, and those few that do would be reregistered. The proposed reform thus would enlarge the public domain, and so reduce the cost of creating new expressive works, at only slight cost to existing copyright owners.

In this example, the law would be allowing copying that had not been specifically authorized by the creator of the intellectual property. Allowing some amount of such copying is a common feature of intellectual property law and one that—for good economic reasons—differentiates it from the law governing physical property. The copyright doctrine of "fair use" already allows a considerable amount of unauthorized copying of copyrighted materials, and Congress, and perhaps the courts as well by interpretation of the copyright statute, could expand "fair use" to

solve problems created by long copyright terms. Judges might rule that the unauthorized copying of copyrighted materials was a fair use if the user had made a reasonable but unsuccessful effort to find and negotiate with the owner of the copyright. Such a rule would create an incentive to establish the kind of registry discussed in the preceding paragraph that would enable copyright owners to protect their copyrights against being deemed abandoned before the copyright term expired. Such a registry would emulate the system by which the costs of transacting in physical property are minimized.

The fair use doctrine has many important applications that involve a weighing of the benefits and costs of allowing a copyright work to be copied (in whole or more commonly in part) without a license from the copyright holder. For example, it has been invoked to authorize parodists to copy large swatches of the parodied work. Courts recognize that a parody is unlikely to be recognizable as such unless there is extensive quotation from the parodied original, and the parodist's costs of transacting with the copyright owner for permission to copy would often be prohibitive, from a social as well as a private standpoint, because the owner would fear that ridicule would diminish the value of the original copyrighted work.

A separate transaction cost argument supports the fair use privilege of quoting short passages from a copyrighted work even if they are not critical of the work, so that a negotiation would not be blocked by the copyright owner's fear of criticism. Although transaction costs in such a case are likely to be low in absolute terms, they are likely to be high relative to the value of the quotation to the copier, while as long as the quotation is only a small fraction of the quoted work, so that the quotation is not a substitute for the quoted work and so will not reduce the demand for it, the cost of the copying to the publisher of the original work will be slight, consisting merely of the loss of a small license fee. Thus, insisting on permission from the copyright owner would harm the copier (and the copier's readers or customers), but do little good for the copyright owner.

A closely related "critic's privilege" permits book reviewers to quote, without the copyright owner's authorization, from the book being reviewed. This is another situation in which, as in the case of a limited copyright term, even creators of intellectual property would probably prefer that free copying be allowed. After all, publishers and authors as a whole benefit from the privilege because it makes book reviews, which are a form of free advertising of books, more credible to the reading public and therefore more valuable to the publishing industry as a whole, though individual publishers may be hurt if many of their books are panned.

Fair use is an example of a doctrine of intellectual property law that has no counterpart in the law of nonintellectual property. There is an economic reason for the difference that is related to the earlier discussion of the intellectual public domain. Generally, only one person at a time can use a given piece of tangible property. It would be inefficient for the law to allow a person to borrow someone else's property without having to negotiate with the owner over terms, because the borrowing imposes a cost on the owner by depriving the owner of the use of the

property, and transaction costs are typically not prohibitive in that setting. But intellectual property is a public good; one person's use does not prevent simultaneous use by another. (An analogy in the physical domain would be a lake used by just a few boaters even at zero access cost, so that the use of the lake by one more boater would impose no costs on the existing users.) If a work is copied, the owner of the original is not prevented from making and selling copies. The existence of cheap copies may inflict economic harm on the original owner, but then again it may not—especially if the copying is limited to excerpts, as in the book review case. In that example and many others, copies are complements rather than substitutes for the original work, and requiring permission would impose a transaction cost with no offsetting benefit.

A more dramatic example of how the public good character of intellectual property alters the economics of misappropriation is the widespread piracy of copyrighted operating system software. This practice is conventionally though imprecisely referred to as a form of “theft”—an example of the pitfalls of too literal an analogizing of intellectual to physical property. While the theft of physical property deprives the property's owner of its use, the “theft” of intellectual property does not. Indeed, if the software pirate, or those who purchase the pirated software from him, could not or would not pay the price charged by the owner of the patented or copyrighted software (this assumes that the owner's ability to price discriminate is limited, so that the owner cannot snag the poorest customers by reducing the price to them to just a shade above the owner's marginal cost), then piracy does not cost the owner any lost sales. Indeed, piracy may increase the income of the original owner if some of the pirate's customers purchase application programs from the owner or if expansion of the owner's user base confers network advantages over competing software producers. However, if piracy reaches a level at which pirates are making copies of their pirated copies and selling the new copies to people who would otherwise buy the copyrighted product, the copyright owner will be hurt, maybe severely.

The public good character of intellectual property and the higher transaction costs of exploiting such property generate the prediction that property rights in intellectual property will be less extensive than in the case of physical property to the extent the law is guided by efficiency concerns. The fair use doctrine is evidence in support of the prediction.

Protecting Expression But Not Ideas

Fair use is only one source of limitations on the scope of copyright; among others is the principle, which has parallels in patent and other areas of intellectual property law as well, that only expression, and not ideas, is protected from unauthorized copying. In the copyright domain, “ideas” are generally not of a scientific, intellectual or practical character, but rather are literary and other artistic tech-

niques and genres, “stock” characters in fiction (such as the wise servant of a high-born foolish master) and basic plot lines. Allowing such building blocks of creative expression to be appropriated would impose enormous transaction costs—no one could write a play, a novel or a book of history without first obtaining a host of licenses—so once again even the creators of intellectual property might prefer that intellectual property protection be limited in this way. In addition, granting exclusive rights to fundamental aesthetic concepts or approaches would induce rent-seeking expenditures on being the first to claim ownership of the new idea. Of course, there are genuine social benefits to discovering a valuable new idea sooner rather than later. But the most fundamental and elementary ideas (like the idea of rhyme or of narrative) would probably have been discovered independently within a short time, so that allowing such ideas to be appropriated would not confer social benefits commensurate with the costs of the race to discover or invent them.

This objection applies to generous patent protection as well—that it will spark patent races that may cost more than the benefits of a slight acceleration in the rate of invention that will result from the race. The objection is unrelated to the traditional concern with patents, however—that, as in the case of copyrights, they limit access by empowering the inventor to charge a price for the use of his invention that exceeds marginal cost, because an invention is information and can often be copied very cheaply unless copying is forbidden. But the objection is related to an objection to expansive intellectual property rights that was noted at the outset of this paper: intellectual property enables pricing above marginal cost, which attracts excessive entry.

The ideas that patent law refuses to protect tend to be ones generated by basic research; they are not considered “useful” within the meaning of patent law. One objection to creating property rights in such ideas (it is also an objection to the protection of expressive “ideas” by copyright) is that it would often be difficult to trace an idea into a particular process or product. Another objection is that patents are a poor method of encouraging basic research, because by definition such research does not have immediate commercial application and hence is not attractive to private investors, especially given the high discount rates of such investors.

Other Alternatives for Covering the Fixed Costs of Innovation

Propertization is not the only method of providing incentives to engage in a socially valuable activity like basic research. Another method, particularly important in scientific fields, is public financing of basic research. It corresponds to the preference of property law for giving finders of lost property a reward rather than the property itself, since giving them the property might overreward finders and thereby induce excessive (rent-seeking) expenditures on finding (and also on not losing!), just as granting patents on the fruits of basic research might cause excessive investment in the inventive process. In the more common case in which

the benefits of basic research are just too remote to interest a firm, public financing of basic research may still be justifiable in economic terms if the social discount rate of such investments is lower than the private rate. In effect, society may weight the welfare of future generations more heavily than private investors do.

A public subsidy for the creation of such property is only one alternative to intellectual property rights. Another, which is feasible if the costs of copying are high, is for government to do nothing because the market will exclude copiers without the aid of the law. Indeed, copyright was not an issue before the invention of the printing press, because of the high cost of copying books by hand. Even today, copyright may not be necessary for unique works of art, such as a painting, since copies of a painting are poor substitutes for the original; that is, the *quality-adjusted* cost of copies is very high.

Sometimes the costs of copying can be *made* high: if an invention can be kept secret at low cost, the inventor will forgo seeking a patent. A patent is limited in duration, requires incurring legal and other fees, and also is a public document that may contain information that allows competitors to “invent around” the patented process. The social as well as private costs of secrecy may be great, but they are to an extent self-limiting; if the invention is a very valuable one, competitors will invest heavily in trying to duplicate it, and so the original inventor may be better off obtaining a patent.

Other alternatives to property-right protection through patent or copyright are a first inventor’s head start, which, if costs of production decline as experience is gained in production, may give the inventor a permanent cost advantage over latecomers; control of a complementary good (such as a search engine for a database the contents of which, being noncopyrightable facts, are not legally protected from duplication); and trademark.

The economic function of a trademark, as noted earlier, is to designate a good or service, so that consumers can pick it out at low cost from an array of competing brands. Almost all the specific doctrines of trademark law flow from this simple consumer-search-cost rationale. For example, trademark protection is forfeited if the trademark is sold “in gross,” that is, separately from the assets used to produce the trademarked product. (“Bulova” could not sell the name for use on another firm’s watch if the Bulova company retained the assets that it used to produce watches. Chiquita Brands, formerly United Fruit Company, could not sell the “Chiquita” banana trademark while retaining the facilities and personnel whereby the company obtains from growers the quality of bananas associated with the trademark.) Unless the trademark remains tied to particular assets, there is no assurance that the trademarked product is the same, from a consumer standpoint, after as it was before the sale; its information value is degraded. To take another example, trademarks are not legally protected if they are “functional” in the sense that a competitor could not make a close substitute for the trademarked product at comparable cost without use of the trademark: an example would be the use of a

wheel as the trademark of a brand of automobile, so that any other automobile sold with wheels would infringe.

A related point is that trademarks lose protection when they become generic, that is, when they designate the product rather than the brand, so that a competitor unable to use the same term to designate his brand could not communicate effectively with consumers. Examples are “yo-yo,” “thermos” and “aspirin,” all of which began as trade names and have become generic. There are analogies to copyright law: the parodist is allowed to borrow extensively from the work being parodied because otherwise he could not create an effective parody, and the doctrine of “merger” of ideas and expression allows the unauthorized copying of copyrighted expression when copying the expression is the only way of conveying the idea.

Many economists used to believe that trademarks were anticompetitive because trademarked brand-name products are sold at higher prices than their generic equivalents. But the belief founders on the difference between the nominal price and the full price of a consumer good. The full price is the sum of the nominal price and the consumer’s search costs, and those costs are lower for branded goods, so that the full price need be no higher than that of a nonbranded substitute. Consistent with this explanation, a trademark is forfeited if the producer fails to maintain the quality of its trademarked good, for in that event the trademark no longer provides the consumer with useful information.

Though a trademark is not itself an intellectual property right, it can be a method of creating or, more commonly, extending such a right. For example, prices of branded pharmaceutical drugs tend to remain high even after the branded drugs come off patent and chemically identical, but much cheaper, generic substitutes appear. If the brand was highly successful in the market, and especially if it had no close substitutes before the patent on it expired, consumers may have become so accustomed to the trade name that even though it hasn’t become generic, they are skeptical that the differently named generic substitutes, though nominally identical, are really as good. Their attitude, which will cause some of them to continue to buy the branded product even though much cheaper generics are now available, and even to pay *more* than before (the producer of the branded product, having lost its price-elastic customers to the generics, may face a less elastic demand from his remaining customers), may be perfectly rational. Heavy investments in trademarks create a hostage situation in which the producer (the hostage) is strongly motivated to maintain quality (which depends on other things besides chemical formula, such as shelf life, quality control and packaging), lest consumer confidence erode. The consumer (the hostage-taker who will “kill” the brand if its quality drops) thereby incurs lower search costs, with a reasonable assurance of the quality of the nominally higher-priced good. The generic manufacturer does not make a comparable investment and so does not provide equivalent assurance.

Given trade secrecy, the learning curve, trademarks and other means of

recovering the fixed costs of invention, the question arises whether patents, which restrict competition further, confer net social benefits. One reason to think that some degree of patent protection is indeed socially beneficial is the heavy social costs of trade secrecy, which would be an even more popular method of internalizing the benefits of invention if there were no patents. These costs are due in part to the fact that licensing trade secrets is a tricky and therefore expensive undertaking because the secret may be revealed inadvertently in the licensing negotiation. A firm that has a trade secret may, for example, decide to do its own manufacturing, even though it has higher costs than other manufacturers, because the cost of licensing the secret to another firm exceeds the difference in the cost of manufacture. Those higher costs of manufacturing are a social cost of trade secrecy and would be avoided by patenting. The trade secret holder may also be reluctant to share its secret with firms in other industries that might employ it more productively than in its original industry, and the productivity loss is again a social cost of trade secrecy. In turn, the costs of trade secrecy provide an economic reason against a prohibition of reverse engineering, since if reverse engineering were prohibited, inventors would have a reduced incentive to seek patent protection in lieu of trade secrecy.

Concern with the monopolistic potential of patents may explain their limited duration, currently 20 years for most types of patent—much shorter than the copyright term. The difference is consistent with the intuition that patents are likely to confer more power over price than copyrights, in part because there is no defense of independent discovery in patent law, as well as with the point noted earlier that long copyright terms reduce a type of congestion and encourage socially valuable maintenance expenditures. The shorter duration of patents may also be related to the legal requirement that the patent application disclose the best means of practicing the patent, even though if the patent is issued no one can practice it without the patentee's authorization. Failure to disclose the best means will invalidate the patent. The disclosure reduces the patentee's power over price by reducing the costs of its competitors. They can use the information disclosed by the patent indirectly, mining it for clues on how to invent around the patent—which is to say how to derive a competitive benefit from the patent without violating the patentee's rights. In effect, just as with copyright law, patent law enables a certain amount of unauthorized exploitation ("fair use") of someone else's intellectual property.

The monopolistic effects of patents are exaggerated in other respects as well. A legal monopoly is not necessarily an economic monopoly; if close substitutes exist for a patented product, the patent may confer little power over price. A common fallacy is to suppose that a patent can be used as a lever to obtain power over the price of unpatented products that are complements of the patented product; in general (though with exceptions discussed in Carlton and Waldman, 2002, and elsewhere), this is false, because increasing the price of a good reduces the demand for its complements. A related fallacy is the idea that a patentee can extend the

length of the patent by demanding that its licensees pay royalties beyond the patent's expiration. The economic value of a patent determines the aggregate royalties that licensees are willing to pay; a patentee who wants to lengthen the stream of royalty payments will have to accept a lower royalty rate.

Given these concerns and tradeoffs, no one knows whether the current scope of patent protection is optimal. It is natural to suppose that a policy of narrowly interpreting patents would be sensible because it would create unequivocal benefits, first in the form of prices closer to marginal cost, and second in terms of a larger public domain, lowering the input costs of later inventors.⁶ On the other side, the costs of narrow patent production are uncertain and perhaps slight. But the question of patent breadth is more complex than this analysis suggests (Gallini, 2002; Mazzoleni and Nelson, 1998; Denicolò, 1996). The narrower a patent, the greater the aggregate costs of the patent system and the more patentees a new inventor may have to obtain licenses from to create a technology that utilizes a broad range of patented knowledge. Conversely, the broader a patent, the greater incentive it creates for the development of truly novel technologies, because merely incremental development will be more likely to infringe an existing patent. Moreover, patents, though sometimes a source of monopoly, are also a counter to monopoly because in their absence firms would have an incentive to expand and combine in order to diversify the risks of invention and internalize the benefits of inventions. A policy of interpreting patents narrowly, which would increase the cost of patent protection to a firm, might swing the balance toward monopoly.

Derivative Works and Improvement Patents

The treatment of derivative works in copyright law differs interestingly from the treatment of their equivalent—improvements—in patent law. A translation or other adaptation or version of an expressive work is a “derivative work” that cannot be copyrighted by anyone other than the owner of the original without the owner's authorization. But the owner can copyright a derivative work only if it differs significantly from the original, as a translation or abridgment does, or the movie version of a book, but not an identical copy in another medium. The requirement that a derivative work have some incremental originality in order to be copyrightable illustrates the sensitivity of intellectual property law to the costs of operating the legal system (Lichtman, 2003). If works indistinguishable from the original were copyrightable, it would be extremely difficult to determine infringement. Suppose *X* copyrights work *A*, and the copyright expires in 1928. In 1927, he copyrights work *B*, which is virtually indistinguishable from *A*. In 1930, *Y* creates work *C*, which is indistinguishable from both *A* and *B*. If *X* sues *Y*, claiming that *Y*

⁶ The extension to the patent area was first made in Scotchmer (1991).

has infringed *X*'s copyright on *B*, and *Y* replies that he was copying *A* (on which copyright had expired) rather than *B*, how could a court decide?

As an alternative to giving the owner of the copyright on the original work the exclusive right to copyright derivative works, the law could allow the creator of a derivative work to copyright it even though he would not be able to sell it without a license from the original owner, since a derivative work invariably copies extensively from the original. Copyright law would then be aligned with patent law, which permits the unauthorized improver of a patented good to patent the improved version, although the improver will not be able to sell it without a license from the original patentee. The patent rule generates "blocking patents": neither the original patentee nor the patentee of the improved version can sell or otherwise use that version without a license from the other. The same regime could be used for copyright works, but is not.

A possible economic justification for the law's different treatment is that technological improvement is typically a continuous, collaborative process, and allowing unauthorized improvers to patent their improvements encourages maximum participation in efforts to improve the originally patented process or product. Progress is much less pronounced in the arts; we do not think that after Shakespeare wrote each of his plays, other playwrights would have been well employed trying to improve them.

This distinction may also explain the different treatment in copyright and patent law of joint ownership. In both domains, a joint owner is allowed to use or license the jointly owned work without the permission of the other owner or owners; this rule reduces transaction costs by eliminating bilateral monopoly. But the joint owner of a copyright who uses or licenses a copyright must account to the other owners for the profits of the use and share them with those others, while the joint owner of a patent need not. The latter rule provides greater encouragement to inventors to keep working to improve their inventions, consistent with the continuously improving quality of technology, but not of the arts.

The patentability of unauthorized improvement patents is supported by the influential "Darwinian" theory of technological progress as a process in which the market selects from among diverse approaches whose relative promise cannot be assessed in advance (Nelson and Winter, 1982). Although a firm has an incentive to license independent inventors of improvements in its patents, its ability to identify the most promising outsiders is likely to be limited by its own knowledge, experiences and firm culture. So there is value in encouraging firms to do research and development on technologies dominated by a patent or patents held by another firm, as the law does by permitting the improvements contributed by those firms to be patented without the dominant firm's authorization.⁷

⁷ The Darwinian theories of innovation cast doubt on Kitch's (1977) influential "prospect theory" of patents, which commends the patent system for centralizing the inventive process in the original "prospector" and thus avoiding parallel inventing. However, the original prospector may have an

An institutional difference between copyright and patent law provides a further explanation for the difference in legal status between derivative works and improvement patents. Copyright is asserted; there is no preliminary administrative screening to determine whether the work sought to be copyrighted is original. In contrast, the screening of patent applications by the Patent and Trademark Office has the consequence that if an improvement patent does not represent a genuine improvement, a patent is unlikely to be issued. Thus, if derivative works were not in the legal control of the owner of the original copyright, a danger would arise of a proliferation of separately owned derivative works that would greatly increase the transaction costs of creators of subsequent derivative works. Suppose *X* copyrights a novel that he has written in English, and *Y* copyrights a French translation. If *Z* wants to do a Spanish translation, he must negotiate with both *X* and *Y*, rather than with *X* alone, since *Z*'s translation is likely to overlap *Y*'s to a degree that would create a serious risk of a suit for infringement if *Z* proceeds without a license from *Y*, as well as, of course, from *X*, since any translation copies much of the original.

This difference between copyright and patent is related to the important legal difference noted earlier that independent discovery infringes patents, but not copyrights. The inventor who unknowingly duplicates an earlier, patented invention is an infringer; a writer who writes a poem that is identical to an already copyrighted poem without his having known of the earlier poem is not. Independent discovery of useful ideas is common, but independent creation of expressive works is uncommon (with the important exception of songs and other short musical works), because copyright protection is limited to the precise expressive form of a copyrighted work. The invention of new products and processes tends, moreover, to be more costly than the creation of new expressive works (motion pictures being a major exception, however). So unless independent discoveries were infringements, inventors would often be reluctant to sink large sums into creating a new product; they would be concerned that the patents they obtained would provide only illusory protection.

Earlier I said there was a question whether copyright was needed for unique works of art, such as paintings, since copies are poor substitutes. Copyright does, however, enable the artist to obtain additional income from derivative works. Hence allowing paintings to be copyrighted increases artists' incomes and presumably therefore the supply of art. That policy might increase economic welfare if works of art were assumed to confer external benefits, though there is little evidence that they do. The broader point is that giving the copyright owner control

incomplete conception of the optimal path of development, in which case centralizing the inventive process may retard technological progress. Notice also that while the objective of the prospect theory is to reduce rent seeking, specifically patent races, by increasing the benefits to being the original prospector, the theory (if implemented) might encourage wasteful races to be the prospector. This argument, by the way, also supports interpreting patents narrowly, since the broader the patent, the more valuable it is and the greater the incentive to race to obtain it.

over derivative works cannot be justified on the ground that it increases the income of copyright owners, unless the supply of copyrighted works is deemed suboptimal.

The Ongoing Expansion of Intellectual Property Rights

The protection of intellectual property rights has expanded in recent decades. One reason is the dramatic fall in the quality-adjusted cost of (and delay in) copying, as a result of digitization; the best known example is computer file sharing of copyrighted music. The lower the cost of copying, the more difficult it is for the owner of the original work to recoup fixed costs in the price the owner charges for copies, unless the owner can prevent competitors from selling copies. Moreover, because of the expanding number and growing wealth of consumers worldwide and the fact that declining costs of disseminating intellectual property have brought more and more of the global market within the reach of producers of such property, the social value of a work that involves a heavy upfront investment, which is the defining characteristic of intellectual property, increases because the incremental cost of providing the work to additional customers is so slight. The social loss from undermining intellectual property rights is greater if, because of cheap copies and lack of intellectual property protection, the incentives to create such works, and hence the quality-adjusted number of such works, are diminished.

Yet reductions in the cost of copying help as well as hurt the owners of copyrighted work, by reducing their marginal costs. The access of intellectual property owners to a global market is greater the lower the costs of disseminating their work. And technological changes that facilitate the direct sale of intellectual property to the ultimate consumer—as when a software producer, bypassing all middlemen, sells and ships directly to consumers over the Internet—enables the owners of intellectual property to bypass intellectual property law by relying on contracts, encryption or both to prevent copying by the consumer—even copying that copyright law would have permitted in the name of fair use or at the expiration of the copyright term. The broader point is that technology and law are substitutes in restricting as well as enlarging access to intellectual property. Copyright owners are naturally most concerned about technological developments, such as music file sharing, that increase access to work they have already created. But other such developments, such as encryption, may reduce access even more—albeit access to the intellectual property of other industries (other than the music industry, for example).

Perhaps most important from a public choice perspective in explaining the expansion of intellectual property rights is the asymmetry of interests between owners of such rights and would-be copiers. Since the owners' principal costs are sunk, almost all the revenue from their sale of copies goes directly to the bottom line, giving them a very large stake in extending their rights. In contrast, would-be copiers, since they will not have exclusive rights once a work is pitched into the

public domain, can expect only a competitive return, and so they have less incentive to challenge intellectual property rights in the legislature than the owners of such rights have to defend them. This asymmetry of interests is probably the reason for the practice of extending copyright and patent terms retroactively, despite the fact that such extensions offer almost no incentive for creating additional intellectual property.

■ *The author wishes to thank James Hines, Timothy Taylor and Michael Waldman for their helpful comments on a previous draft.*

References

- Arrow, Kenneth J.** 1962. "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton: Princeton University Press, pp. 609–26.
- Carlton, Dennis W. and Michael Waldman.** 2002. "The Strategic Use of Tying to Preserve and Create Market Power in Evolving Industries." *RAND Journal of Economics*. 33:2, pp. 194–220.
- Denicolò, Vincenzo.** 1996. "Patent Races and Optimal Patent Breadth and Length." *Journal of Industrial Economics*. 44:3, pp. 249–65.
- Gallini, Nancy T.** 2002. "The Economics of Patents: Lessons from Recent U.S. Patent Reform." *Journal of Economic Perspectives*. 16:2, pp. 131–54.
- Frye, Northrop.** 1957. *Anatomy of Criticism: Four Essays*. Princeton: Princeton University Press.
- Kitch, Edmund W.** 1977. "The Nature and Function of the Patent System." *Journal of Law and Economics*. 20:2, pp. 265–90.
- Landes, William M. and Richard A. Posner.** 2003. *The Economic Structure of Intellectual Property Law*. Cambridge, Mass.: Harvard University Press.
- Lichtman, Douglas.** 2003. "Copyright as a Rule of Evidence." *Duke Law Journal*. 52:4, pp. 683–744.
- Mazzoleni, Robert and Richard R. Nelson.** 1998. "The Benefits and Costs of Strong Patent Protection: A Contribution to the Current Debate." *Research Policy*. 27:3, pp. 273–84.
- Nelson, Richard R. and Sidney G. Winter.** 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: University Press.
- Plant, Arnold.** 1934. "The Economic Theory Concerning Patents for Inventions," in *Selected Economic Essays and Addresses*. London: Routledge & Kegan Paul, pp. 35–56.
- Posner, Richard A.** 2003. *Economic Analysis of Law*. New York: Aspen Publishers.
- Scotchmer, Suzanne.** 1991. "Standing on the Shoulders of Giants: Cumulative Research and Patent Law." *Journal of Economic Perspectives*. 5:1, pp. 29–41.
- Shavell, Steven and Tanguy van Ypersele.** 2001. "Rewards versus Intellectual Property Rights." *Journal of Law and Economics*. 4:2, pp. 525–47.