TOWARD A DOCTRINE OF FAIR USE IN PATENT LAW

MAUREEN A. O’ROURKE


This paper can be downloaded without charge at:

The Boston University School of Law Working Paper Series Index:

The Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract_id=1005650
TOWARD A DOCTRINE OF FAIR USE IN PATENT LAW

MAUREEN A. O’ROURKE

REPRINTED FROM COLUMBIA LAW REVIEW
VOLUME 100, NO. 5 (JUNE 2000)
Copyright © 2000 by Directors of the Columbia Law Review Association, Inc.
TOWARD A DOCTRINE OF FAIR USE
IN PATENT LAW

Maureen A. O’Rourke*

The intellectual property laws are becoming increasingly stressed as their largely time-worn doctrines grapple with problems posed by new technology. In this Article, Professor O’Rourke argues that this pressure has become particularly acute in patent law where policymakers have expanded protection without concomitantly evaluating the impact of that move on follow-on innovation. The traditional assumption that patentees will efficiently license their inventions is breaking down as market failures are becoming endemic. Professor O’Rourke argues that to ensure that patent law achieves its constitutional goals, it should, like copyright law, use a fair use defense to address problems of market failure. Professor O’Rourke proposes just such a defense that, while modeled on copyright doctrine, accounts for and protects patent law’s particular incentive scheme.

INTRODUCTION

The unprecedented domestic economic expansion of the recent past is largely attributable to the success of industries that use and market high technology in its various forms. As the primary law governing rights to the intangible assets that comprise much of the value of firms in these industries, intellectual property statutes have become critical to ensuring continued prosperity. Historically, these statutes have promoted progress through doctrinal constructs designed to fashion an appropriate balance between the grant of exclusive rights to encourage innovation and the maintenance of a viable public domain from which further progress may result. But in the new, fast-paced world of high technology, these time-worn doctrines may simply no longer suffice to perform their traditional function. More particularly, they may be ill-equipped for ensuring that society receives the benefit of new innovations in a timely manner.

* Professor of Law, Boston University School of Law. Thanks to Douglas Baird, Dan Burk, Rochelle Cooper Dreyfuss, Hal Edgar, Ronald Gilson, Mark Lemley, Richard McAdams, Henry Managhan, Ruth Okediji, the Boston University School of Law Library staff, especially Rebecca Troendle Hewitt, Raquel Ortiz, Russ Sweet, and Stefanie Weigmann, and to my research assistants, Mark Bellerman, Aaron Melville, Aaron Feinberg, Anthony Lee, and Bobby Shih. Thanks also to the participants in forums where this paper was presented: Faculty Workshops at Columbia University School of Law and Northwestern University School of Law; seminars at the Engelberg Center on Innovation Law and Policy of New York University School of Law and St. Peter’s College, Oxford University. Special thanks to Bob Bone, Joe Brodley, Peggy Cross, Jane Ginsburg, Wendy Gordon, Keith Hylton, Rob Merges, Mike Moeller, James and Jane Molloy, Liz Porter, and Pam Samuelson, for their comments, suggestions, and time.
Certainly though, the pressure that new technology is placing on the intellectual property system has not gone unnoticed by the relevant public institutions—Congress, the Patent and Trademark Office (PTO), and the courts. But their response to date has primarily been to bring more and more items within the scope of intellectual property protection, without simultaneously re-assessing whether the doctrinal line between the protected and unprotected remains appropriately situated.\(^1\)

This trend is particularly acute in patent law. Indeed, society has entered a "brave new" patent world. While it continues to operate under an "old" statute—one that has not been substantially revised since 1952—the PTO and courts have been construing that statute in an expansive manner.\(^2\) Specifically, in recent years, (i) the judiciary has expanded the subject matter of the patent laws to encompass technologies considered unpatentable in the 1950s as well as others never anticipated fifty years ago;\(^3\) (ii) the PTO has issued patents at a record rate;\(^4\) and (iii) the pri-

\(^1\) The usual academic argument is that rights are "too strong" because Congress, the Patent and Trademark Office (PTO), and the courts have in recent years and to varying degrees tended to "propertize everything." See, e.g., Rochelle Cooper Dreyfuss, Expressive Genericity: Trademarks as Language in the Pepsi Generation, 65 Notre Dame L. Rev. 397, 400-20 (1990) (summarizing how trademark rights have expanded in recent years and why such a trend may be socially undesirable); David Lange, Recognizing the Public Domain, 44 Law & Contemp. Probs., Autumn 1981, at 147, 147 (arguing that "the growth of intellectual property in recent years has been uncontrolled to the point of recklessness" and that "recognition of new intellectual property interests should be offset today by equally deliberate recognition of individual rights in the public domain"); Mark A. Lemley, Romantic Authorship and the Rhetoric of Property, 75 Tex. L. Rev. 873, 898 (1997) ("[T]here is currently a strong tendency to 'propertize' everything in the realm of information. Intellectual property law is expanding on an almost daily basis as new rights are created or existing rights are applied to give intellectual property owners rights that they never would have had in an earlier time."); J. H. Reichman & Pamela Samuelson, Intellectual Property Rights in Data?, 56 Vand. L. Rev. 51, 72-150 (1997) (summarizing the expansion of intellectual property protection regimes in the developed world and discussing the shortcomings of then-pending legislation that would grant "exclusive property rights in the contents of databases").

\(^2\) See infra notes 83-84 and accompanying text (discussing amendments related to the FDA approval process and medical procedure patents), but has not received a major overhaul. In contrast, copyright law was revisited virtually in total in 1976. See William F. Patry, Latman’s The Copyright Law 2-15 (6th ed. 1986) (outlining the history of the Copyright Act).

\(^3\) See, e.g., State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F.3d 1368 (Fed. Cir. 1998), cert. denied, 525 U.S. 1098 (1999) (rejecting the longstanding doctrine holding that business methods are not statutory subject matter under the Patent Act); see also Robert P. Merges, As Many As Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform, 14 Berkeley Tech. L.J. 577, 587 (1999) [hereinafter Merges, Impossible Patents] ("[T]he legal system assumed that intellectual creations were not protectable unless (very) good cause was shown. Today, it often seems the opposite. We now ask: why not protect a new form of intellectual creation?").

\(^4\) See Merges, Impossible Patents, supra note 3, at 601 (graphing the number of patent applications and patents granted in recent years); see also Lawrence Lessig, The Problem with Patents (visited Apr. 23, 1999) <http://www.thestandard.com/article/
mary judicial institution overseeing the system since 1982 (the Federal Circuit) has held patents valid more often than its predecessor courts.\footnote{Display/0,1151,4296,00.html> (stating that the PTO issued 150,000 patents in 1998, a onethird increase over the prior year's total).}

At the same time, however, there has been a sharp increase in the number of contexts in which the grant of a patent license would be socially beneficial, but bargaining for that license would likely break down. Historically, the patent system has assumed that market incentives will, more often than not, lead patentees to exploit their innovations efficiently, often by licensing them to others in the field. Licensing helps to realize patent law's goals both by increasing the probability that the public will benefit from the invention before the patent's expiration, and by encouraging follow-on innovation by licensees and others who purchase the patented product. While market defects may have at times frustrated the conclusion of socially beneficial agreements, until recently, the system functioned in a reasonably efficient manner in the aggregate. Patentees were unlikely to suppress their innovations by refusing to license them, or to use their patents to leverage whatever market power they possessed into secondary markets. Maintaining the strength of the right as an incentive to innovate justified some tolerance toward those refusals to license that did occur, while antitrust law could address anti-competitive conduct intended to expand those rights.

In today's world, however, as the subject matter of the patent law expands, patents proliferate, and high-tech markets evolve, these traditional assumptions may prove incorrect. For example, in the market for operating systems software, which exhibits powerful network effects, strong patent protection can create an insurmountable barrier to entry while also allowing a single patentee to direct innovation in the market for applications running on the dominant system. An operating systems patentee may hold up potential licensees and refuse to participate in group standard-setting efforts that would open the network to the marketplace as a whole, thereby benefiting consumers and advancing innovation. Moreover, the patentee may obtain power sufficient to provide it with leverage over secondary markets well before conventional antitrust causes of action are implicated. By the time they are, it may be difficult to conceive of remedies short of invasive structural relief that might redress the accrued injury to competition.

In other industries, the rapid issuance of patents may result in an opposite but equally damaging dynamic. Essentially, the sheer number of patents creates an "anticommons," where rights are held by so many different patentees that the costs for any one to accumulate all the required licenses to enable production is prohibitive. Yet, for a number of reasons,
including valuation difficulties and strategic bargaining, right-holders may fail to coalesce voluntarily in institutions to pool patents. This, again, challenges the traditional model, which presumes that efficient licensing will occur.

It is therefore a particularly propitious time to re-examine the adequacy of existing law in addressing current conditions. The risk is that, in the absence of a doctrine that permits some privilege of unlicensed use while at the same time protecting patentees' incentives, the patent system may function to thwart the very innovation it is intended to foster.

But reconciling exclusive intellectual property rights with the public interest in furthering innovation is, of course, not a new problem. Virtually since their inception, both the copyright and patent laws have grappled with the question of how to safeguard the incentive inherent in the grant of exclusive rights while at the same time allowing second-comers to build on prior works. Thus, in conducting an evaluation of patent doctrine, it is useful also to look to copyright law as a point of comparison. This analysis reveals a number of differences between the copyright and patent approaches. The most striking and relevant of these for purposes of this Article is the lack of a patent fair use defense comparable to its copyright counterpart.

The copyright doctrine of fair use arose in part, and is justified, as a mechanism to overcome market failures that would otherwise prevent socially desirable uses of the protected work from occurring. For example, courts have always excused some infringements as fair to mitigate problems associated with the anticommons and licensing hold-up. Recently, they have also moved to deploy fair use in addressing issues raised by the network market of software. To the extent that certain market failures are now more commonplace, if indeed not endemic, in the patent system, policymakers should consider adopting a fair use defense.

Such a patent law defense would be a major doctrinal innovation, and one that many, particularly those in the intellectual property field, may find radical, or at least “non-obvious.” Therefore, it warrants a step-by-step exposition, as an unthinking wholesale transplantation of copyright’s fair use doctrine that lacks an appreciation of the existing patent scheme would make little sense. Thus, Part I of this Article considers the salient differences between the copyright and patent schemes, what role a fair use defense would play in patent law, and what that defense might look like. Briefly, the defense would allow courts to weigh defined factors that assess both the social benefit and market harm to the patentee of allowing an infringement to continue. This analysis would help a court to decide both whether to excuse a particular infringement as “fair” and, if so, whether that infringement should continue without compensation to the patentee (as in the copyright fair use model) or whether the infringer should pay the patentee for its continued exploitation of the invention.

In Part II, the Article expands the theoretical analysis and considers likely real-world applications of fair use in patent law. Here, it focuses
first on the particular example of software, drawing again on the copyright experience to illuminate current patent law’s shortcomings. Then it considers other contexts in which a fair use defense might be employed to ensure the appropriate level of innovation.

In Part III, the Article answers at least some of the criticisms likely to be directed at the proposed fair use defense. It argues that fair use is preferable to alternative, piecemeal solutions. The Article also emphasizes that intellectual property law is part of a larger framework that includes other public and private law, including, particularly, antitrust. It concludes that patent fair use can play a helpful role within this system, in part by supplementing and, at times, obviating the need for invasive antitrust enforcement. The Article concludes that fair use, by helping to calibrate exclusive rights in a manner informed by patent policy, is a more desirable, tailored solution to new market conditions than either currently existing patent law or the relatively blunt instrument of antitrust.

I. Why Can’t Patent Be More Like Copyright?: Accounting for the Differences and Arguing for a Fair Use Defense in Patent Law

The patent and copyright statutes have been a part of American law virtually since the inception of the nation. Yet despite their common constitutional basis and shared theoretical underpinnings, their incentive schemes are quite different. To understand why patent law should adopt a fair use defense, it is first necessary to appreciate the common theoretical ancestry of both schemes and their varying implementations of rules and standards consistent with that theory. This comparison illustrates that a copyright-like fair use doctrine in patent law could plug gaps in existing law that are becoming increasingly apparent. Domestic and international authorities offer further support for this argument by noting the desirability of excusing certain infringing conduct, particularly in non-commercial settings. This Article goes beyond these earlier efforts by proposing a detailed fair use defense consistent with patent law’s theoretical underpinnings and flexible enough to excuse even direct infringement conducted with a commercial purpose when appropriate.

A. Patent and Copyright Theory and Implementation

1. Theory. — The theoretical basis of copyright and patent in American law is by now well known. While there are a number of theories that could account for the grant of intellectual property rights, domestic law generally cites an economic rationale for providing intellectual property

6. The first Copyright Act (Act of May 31, 1790, ch. 15, 1 Stat. 124–26) and the first Patent Act (Act of Apr. 10, 1790, ch. 7, 1 Stat. 109–12) were enacted in 1790.
7. See U.S. Const. art. I, § 8, cl. 8 ("The Congress shall have the 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.").
protection in the form of copyright or patent.\textsuperscript{8} Briefly, economic theory supports the grant of intellectual property protection as a response to the public goods problem inherent in the production of information.\textsuperscript{9} In simpler terms, intellectual property protection is a response to the problem of "too little property." Because information is by its nature fluid and can be consumed by more than one person without depleting the amount available to others, once the originator releases the information, he or she cannot easily exclude others from its benefits whether or not they pay.\textsuperscript{10} In the absence of some mechanism to allow the originator to at least recoup his or her investment, information will be underproduced.\textsuperscript{11} One scheme that can function to allow investment recoupment is a grant of exclusive property rights to the originator.\textsuperscript{12} Such a

\textsuperscript{8} However, an economic rationale is by no means the only one that can be offered to justify the grant of intellectual property rights. See Donald S. Chisum et al., Principles of Patent Law 34-50 (1998) (outlining two possible theories supporting patent law: utilitarianism, which focuses on the effect of patent law on public welfare, and a natural rights justification under which "the patent law protects the rights of inventors, regardless of the repercussions on public welfare"); William W. Fisher III, Property and Contract on the Internet, 73 Chi.-Kent L. Rev. 1203, 1212-15 (1998) (citing prominent work and identifying from it four theories of intellectual property: the labor-desert argument based on the writings of John Locke, the utilitarian approach, the self-actualization personality theory, and a new perspective originated by Fisher called proprietorian or social-planning theory, which emphasizes the attainment of a "just and attractive culture"). The point to remember is, as Professor Fisher notes, that the boundaries among different philosophies tend to blur, and the emphasis on one view or another at a particular time tends to change, helping to explain what might seem to be inconsistent caselaw. See Fisher, supra, at 1215; see also John Shepard Wiley, Jr., Copyright at the School of Patent, 58 U. Chi. L. Rev. 119, 140 (1991) (contending that the labor-desert theory and utilitarian perspective "tend to converge as a practical matter, because law governing innovation policy usually must treat creators fairly in order to give them the incentives to act in ways that benefit consumers").


\textsuperscript{11} See Lemley, Improvement, supra note 10, at 995; Wiley, supra note 8, at 138 (brandynd underinvention and underwriting a "standard public goods market failure").

\textsuperscript{12} This is not the only scheme that allows for recoupment. See, e.g., Braga & Fink, supra note 9, at 441 (stating that direct government production, government provision of subsidies and/or government procurement could be alternative arrangements that would encourage the production of knowledge, but arguing that "historical hindsight suggests that market-driven incentives (as exemplified by the proprietary approach) provide the most effective way to organize economic activities, including the creation of knowledge
grant of exclusive rights forms the basis of American intellectual property law in copyright and patent today.

But this description is, of course, deceptively simple. Innovation is by its nature cumulative, building on prior work. The copyright and patent statutes therefore seek not only to reward the first-comer who originates a particular creation, but also to preserve sufficient information for the public to produce additional works. In striking this complex balance, the intellectual property system implements a mixed regime, according some information private property rights and dedicating other information to the commons.\textsuperscript{13} It seeks, in the aggregate,\textsuperscript{14} to guard against both over- and underprotection of information relative to the social optimum\textsuperscript{15} and the concomitant social costs associated with each state.\textsuperscript{16}

\textsuperscript{13} See Carol M. Rose, The Several Futures of Property: Of Cyberspace and Folk Tales, Emission Trades and Ecosystems, 85 Minn. L. Rev. 129, 155 (1998) ("Traditional intellectual property law has generally incorporated some mixture of property and commons, relying on such doctrines as ‘fair use’ in copyright and relying also on the larger climate of First Amendment claims that modify intellectual property law.").

\textsuperscript{14} While it is theoretically possible for the intellectual property system to tailor rights perfectly to reflect the incentive structure of the particular information and industry, such an approach would be prohibitively expensive. The more cost-effective method is that which the system has always employed—implementing an approximation of what the appropriate balance between private property and the commons is. This approach recognizes that mistakes in the form of over- or underprotection of particular items will occur, but aims to achieve the social optimum in the aggregate at a reasonable cost. Note, though, that the statutes do occasionally tailor the grant of rights according to subject matter. See infra note 20.

\textsuperscript{15} The socially optimal level of protection could be defined as that which provides the right owner with just enough incentive to invest in the particular activity while leaving the public with sufficient information from which further progress may result. See Ann Bartow, Educational Fair Use In Copyright: Reclaiming the Right to Photocopy Freely, 60 U. Pitt. L. Rev. 149, 155 (1998) ("To effectuate this bargain [between rightholders and the public], copyright should be no more restrictive than is necessary to create incentives for the promotion of knowledge and learning"); Steven J. Grossman, Experimental Use or Fair Use as A Defense to Patent Infringement, 30 IDEA 243, 255 (1990) ("[T]he rules that govern the exploitation of . . . property need to go no further than to allow the patentee the right to enjoy a substantial share of a patented technology as the price of their labor or risk of investment."); Lydia Pallas Loren, Redefining the Market Failure Approach to Fair Use in an Era of Copyright Permission Systems, 5 J. Intell. Prop. L. 1, 23–24 (1997) ("Of course, in an absolute sense, all that is needed in order to encourage production is the ability to recoup the cost of creation, reproduction, and dissemination."). However, not all would agree with this definition of the social optimum since it proceeds from a utilitarian perspective—another school of thought may offer a different definition. See supra note 8.

\textsuperscript{16} Underprotection results in the problems outlined. See supra text accompanying notes 9–12. If the statutes go "too far" in correcting the public goods problem, they will create problems associated with "too much" property. Overprotection may result in inefficiently skewing resources toward production of the excessively protected type of information. Moreover, overprotection may hamper further progress by making it more difficult for secondcomers to use the protected information.
Both copyright and patent law use a mix of statutory provisions and common law doctrines to calibrate the scope of the respective rights in attempting to implement an appropriate level of protection. In particular, each contains a set of doctrines that limit scope, helping to prevent problems of overprotection in the initial grant of rights as well as thereafter. Comprehending these doctrines requires first an appreciation of the contexts in which they arose.

2. Implementation. — The copyright and patent laws approach the task of correcting the public goods problem quite differently. Copyright law has few requirements for initial protection, but contains a number of doctrines that limit the scope of the right. Patent, on the other hand, while containing some ex post limiting concepts, relies primarily on the PTO’s ex ante evaluation of the invention in determining whether and to what extent to grant a property right. It is relatively less amenable than copyright to adjusting the scope of the right once granted.

Both statutes state threshold requirements for protection. Copyright protects the expression (not the idea) of “original works of authorship fixed in any tangible medium of expression, now known or later devel-

17. For a number of reasons, policymakers should concentrate on preventing overrather than underprotection. While some information likely does suffer from less than the optimal level of exclusive rights under the public law, the industries affected by errors in this direction seem quite capable of correcting the problem, either through the legislative process or the mechanism of private contract. For examples of legislative solutions, see Digital Millennium Copyright Act of 1998, Pub. L. No. 105-304, 112 Stat. 3060 (codified as amended in scattered sections of U.S.C.) (responding to industry fears that electronic technology would unduly undercut the strength of the exclusive copyright rights); 55 U.S.C. § 108 (1994 & Supp. III 1997) (amending the Patent Act to make it easier for biotech patent applicants to overcome obviousness rejections in response to pressure from that industry). For an example of a contractual solution, see ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1453-55 (7th Cir. 1996) (upholding a shrinkwrap license against contract and copyright preemption challenges where the license required an end user to treat non-copyrighted information in a database as if it were copyrighted); see generally Jessica Litman, Copyright and Information Policy, 55 Law and Contemp. Probs., Spring 1992, at 185 (1992) (describing the role of interest groups in forming copyright law); Jessica Litman, Copyright, Compromise and Legislative History, 72 Cornell L. Rev. 857 (1987) (same); J. H. Reichman & Paul F. Uhlir, Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology, 14 Berkeley Tech. L.J. 793, 821-32 (1999) (chronicling bills before Congress addressing database protection). In contrast, the diffuse public that primarily bears the costs of overprotection may be unable to overcome collective action problems to form a coalition to argue for correcting such errors. It relies instead primarily on the courts to safeguard its interests by applying scope limiting doctrines. Finally, at least in the case of patents, the relevant administrative agency, the PTO, may tend, as a matter of institutional structure, to err on the side of granting rather than denying patents to inventions that, in fact, do not meet the statutory standards. See infra note 245 and accompanying text. At the same time, the Federal Circuit, the exclusive Court of Appeals in patent law, has been upholding patents more frequently than its predecessor courts. See supra note 5. However, this Article’s proposal does not depend on a belief that current patent law is systematically skewed toward overprotection. Fair use is a valuable tool to excuse infringements even when the initial grant of protection is perfectly appropriate. See infra Part III.A.
Copyright inhere in eligible subject matter at the moment of fixation, without any need for the creator either to register the work with the Copyright Office or to make it available to the public prior to obtaining his or her rights. Thus, copyright is rather indiscriminate, awarded by operation of law to authors whose works meet the minimal statutory requirements and regardless of whether the public will benefit from disclosure and dissemination of the copyrighted work.

Patent law, in contrast, has exacting threshold standards that an inventor must meet before being entitled to a grant of exclusive rights. To qualify for a patent, an invention must fall within the statutory subject matter, meet the statutory standards of novelty, utility and non-obviousness, and be both disclosed in an enabling disclosure and not

---

18. 17 U.S.C. § 102(a) (1994 & Supp. 1997). A work is “fixed” when its embodiment in a copy or phonorecord, by or under the authority of the author, is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration.” Id. at § 101. Originality is a constitutional requirement, but is not particularly stringent (especially when compared to the patent requirement of novelty), mandating only that “the work [be] independently created by the author (as opposed to copied from other works), and that it possess[ ] at least some minimal degree of creativity.” Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991). Copyright law, in drawing the boundary between itself and patent law as well as between itself and the public domain, expressly excludes protection for “any idea, procedure, process, system, method of operation, concept, principle, or discovery.” 17 U.S.C. § 102(b).

19. See Edmund W. Kitch & Harvey S. Perlman, Intellectual Property and Unfair Competition 536 (5th ed. 1998). Registration is generally required prior to bringing an action for infringement. See 17 U.S.C. § 411. Further, “the certificate of a registration made before or within five years after the first publication of the work shall constitute prima facie evidence of the validity of the copyright and of the facts stated in the certificate.” Id. § 410(c). The Copyright Office engages in a variety of tasks other than taking registrations, including advising Congress on legislative changes. For a description of the Office’s role and mission, see United States Copyright Office: A Brief History and Overview (visited Apr. 20, 2000) <http://www.loc.gov/copyright/docs/circ1a.html>.

20. The particular bundle of rights varies depending on the nature of the subject matter. See, e.g., 17 U.S.C. § 114(a) (“The exclusive rights of the owner of copyright in a sound recording do not include any right of performance under section 106(4).”). The Copyright Act also provides for compulsory licenses in certain circumstances. See infra note 35.


22. See id. § 102.


24. See 35 U.S.C. § 103 (stating that a patent will not issue for an invention if the differences between it and the prior art “are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which such subject matter pertains”).

25. See id. § 112 (requiring a description “to enable any person skilled in the art . . . to make and use” the invention).
otherwise barred from protection by the statute. A PTO examiner evaluates an inventor’s patent application for compliance with these standards and negotiates with the inventor over the proper wording of the patent’s claims. These claims, in turn, form the metes and bounds of the inventor’s property right—a right the inventor does not obtain until the PTO issues the patent. The quid pro quo for the grant of a patent is the statutorily-mandated disclosure that adds to the store of public knowledge and the dedication of the invention to the public once the patent’s term expires.

Given these considerably different threshold requirements for protection, it is not surprising that patent rights, while substantially temporally shorter than copyright rights (date of filing plus twenty years versus life of the author plus seventy), are more extensive. A copyright primarily grants its owner the exclusive right to reproduce the work in copies, while the patent allows the patentee to prevent any other from making, using, or selling the invention (or its equivalent) in the United States. Moreover, under copyright law, independent creation is a defense to infringement, but under patent law it is not.

26. See id. § 102 (describing conditions under which an inventor may lose an entitlement to a patent including if the invention was placed on sale more than one year before the date of the patent application).


28. See Marsh v. Nichols, Shepard & Co., 128 U.S. 605, 612 (1888) (“Until the patent is issued there is no property right in it, that is, no such right as the inventor can enforce.”).

29. See Beidler v. United States, 233 U.S. 447, 453 (1914) (“A correct and adequate description or disclosure of a claimed discovery... is essential to the validity of a patent, for the reason that such a disclosure is necessary in order to give the public the benefit of the invention after the patent shall expire. The source of the power to grant patents, and the consideration for granting them, is the advantage which the public will derive from them, especially after the expiration of the patent monopoly, when the discoveries embodied in them shall become a part of the public stock of knowledge.”) (citation omitted).


33. This is implicit in the respective statutory grants of exclusive rights. See generally Tj, Inc. v. GMA Accessories, Inc., 132 F.3d 1167, 1169 (7th Cir. 1997) (“The Copyright Act forbids only copying, if independent creation results in an identical work, the creator of that work is free to sell it.”). The evidentiary weight of administrative action also varies under the two statutes. While a copyright registration certificate is prima facie evidence of the copyright’s validity, see supra note 19, a duly issued “patent shall be presumed valid.” 35 U.S.C. § 282.
The two systems also each employ scope-limiting doctrines to guard against overprotection. A brief review of these devices reveals not only that copyright law is more tolerant of a certain amount of infringement than patent, but also that this tolerance is not simply a logical by-product of copyright’s relatively low investment in evaluating whether a work merits protection. Rather, it performs socially useful functions that patent law, even with its substantial upfront investment in making the protection decision, should find it desirable to incorporate.

B. Defining the Scope of the Exclusionary Right—Copyright’s Fair Use Doctrine and Its Patent “Equivalents”

Because tailoring rights by type of work at the statutory level can be quite expensive, Congress has often relied on vaguely defined standards rather than bright-line rules in drafting both the Copyright and Patent Acts.34 Courts construe both these standards and common law doctrines to define more sharply the particular right’s scope. In copyright law, they have excused infringement as fair often in circumstances where market failures would otherwise render the exclusive rights overbroad and prevent socially efficient and desirable uses of the copyrighted work from occurring. Patent law contains no directly analogous doctrine, yet the same market defects with which copyright fair use grapples are increasingly arising in markets for patented products. A patent fair use defense that explicitly balances the patentee’s exclusive rights against overall social welfare would better equip the law to address these defects.

1. Copyright Law’s Scope-Limiting Doctrines—Including Fair Use. —
Within the Copyright Act, the most important sections limiting the scope of exclusive rights are the prohibition against granting copyright protection to ideas35 and the provision extending a fair use defense to infringers in certain circumstances.36 The former operates to draw the line between copyright and patent, as well as between copyright and the public domain, by denying copyright protection to ideas.37

34. However, some tailoring does occur. For example, the Copyright Act grants different rights for certain subject matters. See supra note 20. See generally J. H. Reichman, Legal Hybrids Between the Patent and Copyright Paradigms, 94 Colum. L. Rev. 2432 (1994) (describing and decrying a trend among many countries to enact technology-specific protective legislation).

35. See 17 U.S.C. § 102(b). For other examples of scope limitations, see, e.g., id. § 109(a) (under the “first sale” doctrine, the copyright owner cannot use its public distribution right to control re-sales conducted by purchasers of the copyrighted work); id. § 115 (requiring compulsory licensing for “making and distributing phonorecords”).

36. See id. § 107.

37. Yet the dividing line between idea and expression is not simple to draw. The seminal case is Baker v. Selden, 101 U.S. 99 (1879), in which the Court addressed the distinction between a book that can be protected by copyright and the system or art described by that book, that can only be protected, if at all, by patent. See id. at 104–05; see also Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) (“[N]o principle can be stated as to when an inquirer has gone beyond copying the [unprotected] ‘idea,’ and has borrowed its [copyrighted] ‘expression.’ Decisions must
In contrast, the fair use defense does not operate as a blanket rule affording or denying copyright protection to particular information. Rather, it is a long-standing equitable doctrine that fine-tunes the scope of a copyright over time. Fair use has traditionally been defined as a "privilege in others than the owner of the copyright to use the copyright material in a reasonable manner without his consent." In effect, a successful fair use defense imposes a limited royalty-free compulsory license on the copyright owner: The party asserting the defense has infringed, but that infringement is excused. What theory supports such a limit on the copyright owner's rights?

Generally, the cases reveal that market failure may sometimes justify labeling an infringing use as fair if it is socially desirable and excusing it will not substantially harm the copyright owner's incentives. These market failures generally fall into one of three categories: (i) high transaction costs that frustrate private bargaining; (ii) positive externalities that prevent the infringer from being able to pay the copyright owner's price for a license; and (iii) the failure of any market for the particular use to develop. Some brief examples illustrate both the types of cases in which such failures tend to occur and why excusing infringement as fair under the particular circumstances is socially desirable.

In Sony Corp. of America v. Universal City Studios, Inc., the Supreme Court refused to enjoin Sony from marketing VCRs that consumers used to make unauthorized copies of copyrighted broadcasts, holding the
copying excused under the fair use doctrine.\textsuperscript{44} That consumers would face insurmountable transaction costs in identifying, contacting and contracting with the individual copyright owners for permission to tape helps to account for the holding.\textsuperscript{45} Moreover, the impact of home copying on the copyright owners’ market was unlikely to impact negatively authors’ incentives to invest in new creations.

In yet other contexts, often those involving criticism or commentary, the market failure is not high transaction costs, but the existence of positive externalities.\textsuperscript{46} For example, a critic may wish to use a copyright owner’s work to write a trenchant social commentary. However, he or she may be unable to pay the price of permission because the market structure will not allow the critic to reap revenue from "persons who might gain knowledge from the public debate sparked by the book without having purchased the book itself."\textsuperscript{47} A fair use defense might be justified in such cases if the public benefit is large and the impact on the copyright owner small.

In other cases, particularly those involving parody, the market failure may be the refusal of the copyright owner to grant a license at any price.\textsuperscript{48} In certain markets, then, the "usual economic assumption that the owner of a resource will either exploit that resource himself, or will sell it to someone else who will"\textsuperscript{49} may not hold. As the Supreme Court says, "the unlikelihood that creators of imaginative works will license critical reviews or lampoons of their own productions removes such uses from the very notion of a potential licensing market."\textsuperscript{50}

\begin{footnotesize}
\textsuperscript{44} See id. at 454–55.
\textsuperscript{45} See Gordon, supra note 42, at 1655. When transaction costs decline, the scope of fair use may also contract. See, e.g., American Geophysical Union v. Texaco Inc., 60 F.3d 913, 931 (2d Cir. 1994) ("[I]f it is sensible that a particular unauthorized use should be considered 'more fair' when there is no ready market or means to pay for the use, while such an unauthorized use should be considered 'less fair' when there is a ready market or means to pay for the use.").
\textsuperscript{46} See Gordon, supra note 42, at 1630–35 (identifying market failures in addition to high transaction costs, including the existence of positive externalities borne by the public, the problem of assigning monetary value to social values, and anti-dissemination motives that undercut the usual assumption that rightholders will efficiently license their rights); Loren, supra note 15, at 48–53 (describing situations where external benefits may not be internalized, and information will be underproduced to society’s detriment).
\textsuperscript{47} Gordon, supra note 42, at 1630 (giving criticism as an example of a case in which the inability of the critic to internalize positive social benefits may lead to market failure).
\textsuperscript{49} Gordon, supra note 42, at 1632.
\textsuperscript{50} Acuff-Rose, 510 U.S. at 592.
\end{footnotesize}
However, courts should be reluctant to second-guess refusals to license: "Market failure should be found only when the defendant can prove that the copyright owner would refuse to license out of a desire unrelated to the goals of copyright, notably, a desire to keep certain information from the public"—a so-called "antidisemination motive" that, if considered legally cognizable, would undercut copyright law's goals. Thus, it might also be termed an "anti-copyright" motive. Fair use under this model often involves strong First Amendment interests in public discourse, helping to ensure that the public benefit exceeds the cost to the rightholder.

The policy arguments that support allowing certain uses to proceed as "fair" also identify fair use as a tool that judges can use to refine a copyright's scope to prevent it from becoming overbroad. However, these arguments do not answer the question of whether the use should proceed for free. Historically, fair use has been an all-or-nothing defense: if successful, the infringer may continue the infringing behavior without payment to the copyright owner; if unsuccessful, the infringer will likely be enjoined from continuing his or her conduct and ordered to pay damages.

Unfortunately, the cases do not separately address the questions of the propriety of the infringing use and the failure to pay for it, collapsing the logic into one inquiry—perhaps because the same considerations that excuse the use may also be a factor in excusing the failure to pay for it. For example, it may be "good" policy to excuse payment if it is simply too impractical for the copyright owner to devise a payment scheme or, obvi-

51. Gordon, supra note 42, at 1634.

52. See, e.g., Jay Dratler, Jr., Distilling the Witches' Brew of Fair Use in Copyright Law, 43 U. Miami L. Rev. 233, 235 (1989) (noting that the "flexibility of fair use allows judges to adjust the contours of copyright protection to fit individual circumstances as they arise, in the true common law tradition"); Loven, supra note 15, at 21 (identifying fair use as "a critical component of copyright law that keeps the monopoly of copyright from becoming too broad").

53. See Memo from Wendy J. Gordon, Aug. 25, 1999 [hereinafter Gordon, Memo] (on file with the Columbia Law Review) ("[E]xisting case law [ ] views 'fair use' as an all-or-nothing choice between no liability, or liability plus injunction."). Judge Alex Kozinski has suggested that Congress discard both fair use and injunctive remedies and focus instead on calibrating remedies appropriately, allowing certain infringing uses to proceed for a fee. See Alex Kozinski, What's So Fair About Fair Use?, Brace Memorial Lecture 1–10 (Nov. 11, 1999) (transcript on file with the Columbia Law Review). This Article does not go so far. Patentees would still be entitled to an injunction unless the proposed fair use test indicates that society would benefit from allowing the use to continue. Courts then must decide whether the use should proceed for free or not. The difference between this Article's approach and that of Kozinski is at least partly attributable to the fact that patented inventions generally lack the expressive elements (and the accompanying First Amendment concerns) of copyrighted works. But see Dan L. Burk, Software as Speech, 8 Seton Hall Const. L.J. 683, 690–91 (1998) [hereinafter Burk, Software as Speech] (suggesting that patented software may implicate First Amendment issues because of the convergence between patents and copyrights).
ously, if requiring payment would prevent the use from occurring.\textsuperscript{54} In other cases, the infringer's use may simply be of a type that policymakers do not want to commodify, often because it seems normatively rightful that the infringer not have to pay for a particular use, such as criticism or the response thereto.\textsuperscript{55}

Nevertheless, the Supreme Court has hinted that there may be cases in which the large social benefit associated with an infringing "unfair" use would justify a court in awarding damages to the copyright owner, but refusing to grant what is the routine remedy in intellectual property infringement cases—an injunction.\textsuperscript{56} This leaves open the question of whether there might also be cases in which the fair user, although allowed to continue its infringement, should pay for it.

Although theory supports granting fair use rights in certain circumstances, policymakers obviously must give the doctrine some coherence to avoid both impairing incentives by excusing too many infringements and harming the public by excusing too few. Section 107 of the Copyright Act contains examples of the types of uses often considered to be "fair," as well as four non-exclusive factors that courts may consider in evaluating a particular use.\textsuperscript{57} These factors focus primarily on market harm, leading courts to be more receptive to a fair use defense where the infringer has added substantial value to the original work and "transformed" it in some way.\textsuperscript{58} However, the inquiry is heavily fact-intensive, with no one factor determinative.

\textsuperscript{54} See Gordon Memo, supra note 53; see also supra note 47 and accompanying text (explaining that where an infringer cannot capture the positive externalities associated with a use, he or she may not be able to pay the copyright owner to continue the infringing use).

\textsuperscript{55} See Gordon, Memo, supra note 53 ("Failing to pay is justified if we DON'T WANT the users to pay, because the resource is one that shouldn't be commodified, or the user's behavior is of a sort we don't want to place in the copyright owner's control."); see also Maxtone-Graham v. Burtchaell, 803 F.2d 1233 (2d Cir. 1986) (allowing a Catholic priest to quote from a pro-choice book in writing his own critique of interviews of "abortion veterans").

\textsuperscript{56} See Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 n.10 (1994) ("Because the fair use enquiry often requires close questions of judgment as to the extent of permissible borrowing in cases involving parodies (or other critical works), courts may also wish to bear in mind that the goals of the copyright law, to stimulate the creation and publication of edifying matter, ... are not always best served by automatically granting injunctive relief when parodists are found to have gone beyond the bounds of fair use."); see also American Geophysical Union v. Texaco, Inc., 60 F.3d 913, 932, n.19 (2d Cir. 1994) ("[W]e note that the context of this dispute appears to make ill-advised an injunction, which, in any event, has not been sought. If the dispute is not now settled, this appears to be an appropriate case for exploration of the possibility of a court-imposed compulsory license.").

\textsuperscript{57} See 17 U.S.C. § 107 (1994 & Supp. III 1997); see also infra note 118 (setting forth the four factors).

\textsuperscript{58} See Acuff-Rose, 510 U.S. at 579: Although [a] transformative use is not absolutely necessary for a finding of fair use . . . the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works. Such works thus lie at the
down because the parties hold different assumptions about a key negotiating point like the value of the other's technology. Strategic bargaining based on erroneous assumptions may result in an impasse, imposing large costs on the public by denying it the benefit of the new technology. The reverse doctrine of equivalents may provide the infringer with a negotiation threat credible enough to increase the probability that the parties will conclude a licensing agreement.

At first glance, the doctrine of blocking patents picks up where the reverse doctrine of equivalents leaves off. "Blocking patents" refers to the situation in which an inventor obtains a patent on an improvement of a previously patented invention. This inventor infringes the original patent, but the first patentee has no rights to the patented improvement. Because the first patentee usually has an incentive to obtain access to the improvement, the blocking patent gives the infringer some bargaining power in negotiations. By helping to overcome market defects ranging from high transaction costs to strategic behavior in negotiations, the blocking patents doctrine should increase the probability that the parties will reach a mutually beneficial licensing agreement.

The judicially created experimental use exception may also assist a small number of infringers. Under this doctrine, courts may excuse infringement when the infringer's use is for non-commercial, experimental purposes. Unlike copyright fair use, experimental use will not shelter the infringer who reverse engineers the patented invention, invents around it, and offers a new, non-infringing product for sale. Indeed, precedent suggests that the exception will not apply if the infringer has a commercial motive, regardless of how remote the infringement is from realization of that motive.

---


66. See id. at 91 (noting that past government interventions to force licensing indicate that such breakdown does occur with some frequency).

67. See Lemley, Improvement, supra note 10, at 1099–1110.

68. See id. at 1062; see also id. at 1065 ("[O]ne justification for the blocking patents doctrine [is that] it may encourage efficient licensing transactions in situations where they otherwise would not occur due to market imperfections.")

69. See Rebecca S. Eisenberg, Proprietary Rights and the Norms of Science in Biotechnology Research, 97 Yale L.J. 177, 220–26 (1987) [hereinafter Eisenberg, Proprietary Rights] (describing the evolution of the doctrine in caselaw and suggesting how it might be used to help encourage research in the biotechnology field).

70. See Roche Prod. v. Bolar Pharmaceutical Co., 733 F.2d 858, 858–63 (Fed. Cir. 1984) (stating that the defense does not permit "unlicensed experiments conducted with a view to the adaptation of the patented invention to the experimenter's business," as opposed to experiments conducted "for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry"); Rebecca S. Eisenberg, Patents and the Progress of Science: Exclusive Rights and Experimental Use, 56 U. Chi. L. Rev. 1077, 1092 (1989) [hereinafter Eisenberg, Exclusive Rights].
Finally, patent law limits scope through application of its misuse doctrine which pre-dates that of copyright law and on which, in fact, copyright’s misuse defense is modeled. Patent misuse occurs when the patentee has “impermissibly broadened the ‘physical or temporal scope’ of the patent grant with anti-competitive effect.” Patentee practices generally fall into one of three categories: (i) those like extending the royalty term beyond the patent’s expiration that constitute per se misuse; (ii) those that are statutorily protected under Section 271(d) of the Patent Act from categorization as misuse; and (iii) all others, assessed under a rule of reason analysis. Although amendments to the Patent Act in recent years have given patentees some relief from the doctrine, it persists, separate and apart from antitrust law, although how much so is a matter of continuing debate.

71. See, e.g., Lasercomb Am. Inc. v. Reynolds, 911 F.2d 970, 975–77 (4th Cir. 1990) (describing the evolution of patent misuse and explaining that, because copyright has analogous goals, it is appropriate to adopt a misuse defense in copyright as well). 72. Virginia Panel Corp. v. MAC Panel Co., 133 F.3d 860, 868 (Fed. Cir. 1997) (quoting Windsurfing Int'l, Inc. v. AMF, Inc., 782 F.2d 995, 1001 (Fed. Cir. 1986) (quoting 3 Wonder-Tongue Lab., Inc. v. University of Ill. Found., 402 U.S. 313, 345 (1971))). 73. See Brulotte v. Thys Co., 379 U.S. 29, 32 (1964) ("[A] patentee’s use of a royalty agreement that projects beyond the expiration date of the patent is unlawful per se."). 74. See 35 U.S.C. § 271(d) (providing that a patentee is not guilty of misuse for (i) deriving revenue from or licensing another to engage in acts that would constitute contributory infringement in the absence of the patentee’s consent; (ii) enforcing the patent; (iii) refusing to license the patent; or (iv) tying a license to acquisition of another license or product unless the patentee possesses market power). 75. See Patricia A. Martone et al., The Patent Misuse Defense—Does it Still Have Vitality?, 524 PLI/Pat. 289, 298 (1998). The rule of reason analysis evaluates the totality of the circumstances, including the nature of the business and the restraint and its effect. See id. (noting that the rule of reason inquiry takes into account a variety of factors, including specific information about the relevant business, its condition before and after the restraint was imposed, and the restraint’s history, nature and effect); see also Virginia Panel Corp., 133 F.3d at 869 (stating that a court must determine if the challenged conduct is reasonably within the patent grant, i.e., that it relates to subject matter within the scope of the patent claims) (quoting Mallinckrodt, Inc. v. Medipart, Inc., 970 F.2d 700, 708 (Fed. Cir. 1992)). 76. In 1988, Congress adopted the Patent Misuse Reform Act, Pub. L. No. 100-703, 102 Stat. 4676 (1988), which added subsections (4) and (5) to 35 U.S.C. § 271(d), further limiting the defense of patent misuse. 77. See Transparent-Wrap Mach. Corp. v. Stokes & Smith Co., 329 U.S. 637, 641 (1947) ("The requirement that a licensee under a patent use an unpatented material or device with the patent might violate the antitrust laws. . . . The condemnation of the practice [on misuse grounds], however, does not depend on such a showing."); Robert P. Merges, Who Owns the Charles River Bridge? Intellectual Property and Competition in the Software Industry 27–28 (unpublished manuscript on file with the Columbia Law Review) (hereinafter Merges, Charles River Bridge) (explaining that patent misuse remains an “odd but stubborn variant of antitrust law,” separate and apart from it, despite commentators’ objections to the vagueness of the doctrine); Note, Is the Patent Misuse Doctrine Obsolete?, 110 Harv. L. Rev. 1922, 1927–28 (1997) (comparing patent misuse to antitrust and concluding that there is some role for misuse that differentiates it from antitrust).
Thus, both copyright and patent law use a mixture of statutory provisions and common law principles to define the scope of the rightholder’s intellectual property. But the nature and content of these definitional doctrines varies. The following argues that new conditions justify a bit of doctrinal convergence across the two systems under the rubric of fair use.

C. Accounting for The Differences—Why Patent Fair Use?

Commentators have considered why the copyright and patent laws use varying incentive structures. Their suggestions have served primarily to illustrate that there is no ultimate universal principle compelling copyright and patent laws to have different incentive schemes. Indeed, history reveals that each set of law has freely borrowed doctrine from the other when appropriate. For example, copyright law modeled its misuse defense on that of patent, and both schemes employ a “first sale” doctrine that enables purchasers of protected items to use and resell them without fear of being held infringers. Why would patent law not find it equally useful to borrow the concept of fair use from copyright?

At first glance, the answer seems to be that patent’s higher threshold requirements in conjunction with existing scope limiting doctrines essentially function like copyright fair use. In particular, the reverse doctrine

78. See, e.g., Burk, Patenting Speech, supra note 40, at 67 (noting the argument that because patented technology is more expensive to develop it requires more of an incentive for its development); Paul Goldstein, Infringement of Copyright in Computer Programs, 47 U. Pitt. L. Rev. 1119, 1121 (1986) (stating that patent’s high standards encourage investment only in substantial technological developments, while copyright’s minimal requirements maximize the quantity of expression); Wiley, supra note 8, at 182 (arguing that examination prior to the grant of a copyright would involve “paralyzing costs” because there are so many copyrighted works and claiming that the “different characters of the innovative processes the two regimes protect” help to explain certain variations between the two); Douglas Y’Barbo, The Heart of the Matter: The Property Right Conferring the Copyright, 49 Mercer L. Rev. 643, 665–66 (1998) (arguing that “genuine differences in the attributes of the property that each regime is designed to protect,” including copyright’s protection of the aesthetic versus patent’s protection of the functional, helps to explain the differences between the two regimes of protection). But see Rochelle Cooper Dreyfuss, A Wise Guy’s Approach to Information Products: Muscling Copyright and Patent into a Unitary Theory of Intellectual Property, 1992 Sup. Ct. Rev. 195, 222 (rejecting the argument that copyright and patent differ because they address the needs of different industries); Lemley, Improvement, supra note 10, at 1038 n.239 (rejecting Wiley’s volume argument as circular). None of these rationales can really explain differences in the systems beyond the varying threshold standards for protection. Implicit in at least some of the authorities is the belief that patented inventions are more likely to confer market power on the rightholder than ownership of a copyright. Therefore, the system should scrutinize the award of a patent more closely. This too is unsatisfying. Market power is as much an effect as it is a cause of the different statutory schemes. If patent were to grant less extensive rights, patentees would be less likely to obtain market power. Further, as the Microsoft case illustrates, copyrights can enable market power as much as patents. See United States v. Microsoft Corp., 8 F. Supp. 2d 30, (D.D.C. 2009) (holding that Microsoft illegally maintained monopoly power in the operating systems market and noting that copyright law could not justify the restrictive contractual provisions at issue).

79. See supra note 62 (explaining “first sale” doctrine); supra note 71.
of equivalents and experimental use doctrine seem quite similar to copyright fair use.

Additionally, the theoretical justification for copyright fair use—the "market failure" rationale—historically has seldom been implicated in patent law. For example, patent licensing has generally proceeded in an environment of transaction costs that, while not low in an absolute sense, have been relatively low enough for a vibrant patent licensing market to exist. The reverse doctrine of equivalents and doctrine of blocking patents stand ready to provide an incentive to license to those who might otherwise not. Also, in those cases in which barriers such as high transaction costs have frustrated licensing despite these scope-limiting doctrines, private parties have solved the problem by pooling patents or setting up more or less formal institutions to facilitate the exchange of rights.

Likewise, the public benefit approach—which justified copyright fair use as a way to allow socially beneficial uses that generate large positive externalities—is at least indirectly addressed by existing patent doctrine. By sheltering the radical improver, the reverse doctrine of equivalents helps to ensure that the public will receive the benefit of major advances. The blocking patents doctrine increases the probability that the public will profit from substantial improvements. Those minor improvers unable to obtain a patent will be held infringers. But the cost of keeping a minor advance from the public may be justified to maintain the original patentee’s incentive.

Moreover, Congress has enacted tailored provisions to address specific situations in which the public interest demands some weakening of the patent right. For example, the Patent Act allows certain infringements when they are a step in obtaining Federal Drug Administration approval for a new drug. It also limits the ability of patentees to recover damages for infringements that occur when a medical practitioner infringes a patented medical or surgical procedure while treating a pa-

80. See supra notes 41–51 and accompanying text (discussing the rationale supporting fair use).

81. See Merges, Blocking Patents, supra note 65, at 78 (explaining patent law as a system of property rules justified under the Calabresi and Melamed framework in part by relatively low transaction costs, and noting that the patent licensing market is large and thriving, indicating parties’ abilities to engage in private transactions).


tient. Both examples reflect situations in which the public benefit from the infringement may be so great that it outweighs the patentee’s interest in its exclusive rights.

Further, patented inventions are simply not imbued with the same First Amendment interests that copyrighted material tends to be, making it less likely that a fair use approach to excuse infringement would be warranted. To the extent that a patentee has an anti-dissemination motive like that often used to justify the grant of fair use in First Amendment cases, the doctrine of patent misuse may help to thwart it—at least in part—by policing overly restrictive licensing terms.

What then would be the role for patent fair use? Are there cases of market failure with which patent doctrine does not adequately deal? The following section argues that there are such cases. It begins by drawing support from both domestic scholars and international movements that indicate an increasing appreciation of the desirability of permitting certain infringements that would not be excused under current doctrine to continue. By identifying gaps in existing law, these authorities assist in formulating a patent fair use principle. However, they fall short of setting forth a comprehensive fair use doctrine that transcends particular contexts.

D. Toward a Doctrine of Fair Use in Patent Law

Scholars have long examined the patent system with an eye toward proposing cost-effective improvements. However, many of these suggestions focus more on careful application of, or minor adjustments to, existing doctrine rather than incorporation of new concepts in substantive patent law. But the time is now to move beyond earlier authorities and adopt a doctrine of fair use that brings the balance between exclusive rights and the public welfare that implicitly informs conventional doctrine into the open. A fair use doctrine will give courts defined factors to weigh that will prevent them from having to contort existing law to arrive at desirable results, and better ensure that the patent system is faithful to its constitutional goal.

1. The Domestic View. — In groundbreaking work, Professor Rebecca Eisenberg considered the impact of strong patent rights on scientific re-


85. However, the Sixth Circuit recently held that source code is protected by the First Amendment. See Junger v. Daley, No. 98-4045, 2000 WL 343566, at *4 (6th Cir. Apr. 4, 2000). The Ninth Circuit had earlier held similarly but withdrew its opinion for rehearing. See Bernstein v. United States Dept. of Justice, 192 F.3d 1308 (9th Cir. 1999); see also Burk, Patenting Speech, supra note 40, at 63–76 (discussing whether some adjustment to patent doctrine, potentially including the addition of a fair use defense, is required to address free speech issues raised by patented software); Burk, Software as Speech, supra note 55 (noting that patent law was not written to accommodate First Amendment interests in the same way that copyright law, with its fair use doctrine, was).
search. She suggested that a researcher who infringes a patent in the course of verifying the functionality of the patented invention be exempt from infringement liability. As she stated, patentees are unlikely to subject their inventions to scrutiny, except to "like-minded colleagues who are predisposed to accept their claims and who will not challenge their fundamental assumptions." But a critical examination is important in maintaining social control within the scientific community and provides the basis for further progress. In copyright fair use jargon, the patentee may have an anti-dissemination motive (here, anti-patent) that would prevent an otherwise socially beneficial use from occurring, frustrating the patent system's goals. According to Professor Eisenberg, patent's experimental use exception could be tweaked to achieve the desirable result.

Furthermore, she argues that a patentee should not be granted an injunction against research use of the patented invention that could lead either to improvements of or alternatives to the invention. In both cases, high transaction costs, valuation uncertainties, and the patentee's desire to maintain its monopoly may prevent the parties from concluding a license. The inability of the infringer to internalize the diffuse social benefit of scientific progress will render it unable to pay the price the licensor demands—if it is willing to negotiate at all. But if the research were enjoined, the public would be deprived of continuing scientific progress. However, Professor Eisenberg does not suggest a fair use defense. Rather, she argues that where research eventually leads to a patented improvement, the original patentee's incentives will be protected by the doctrine of blocking patents; where the research leads to a non-infringing substitute product, those incentives could be protected by a modified experimental use exception that allows a court to award a reasonable royalty for the research infringement.

86. See generally Eisenberg, Exclusive Rights, supra note 70, at 1066–78; Eisenberg, Proprietary Rights, supra note 69, at 199–203.

87. See Eisenberg, Exclusive Rights, supra note 70, at 1078 ("Research use of a patented invention to check the adequacy of the specification and the validity of the patent holder's claims about the invention should be exempt from infringement liability.").

88. Id. at 1055.

89. See id. at 1065–66 (noting that this examination helps to establish the value of contributions and that free access, by lowering research costs, fosters further progress, and stating: "Free access promotes scientific progress by fostering independence among scientists, allowing individual researchers to work on the problems they choose in the manner they see fit without needing to obtain approval from other scientists who may disagree with their approaches, or having to disclose their plans to their research rivals.

90. See id. at 1074.

91. See id. at 1078.

92. See id. at 1075–74.

93. See id. at 1077–78.

Damages for the research use may be unnecessary if the original patent is broad enough in scope to cover the improved technology developed by the researcher. In this situation, the patent holder's interests will be adequately protected by
In a similar vein, Stephen Grossman assessed the consequences of the proliferation of patents held by non-profit institutions for scientific progress. He notes that one result of this phenomenon may be an increase in the number of instances in which a patentee seeks damages for infringement but the infringing conduct itself (like teaching or research) is actually consonant with realizing the purposes of the patent system: The infringer advances research without harming the patentee. However, he does not advocate that patent law adopt a fair use defense. Rather, he argues that patent courts, in deciding infringement claims, have historically considered a number of factors—including the infringer’s intent, the economic harm to the patentee, and whether the infringement is de minimis. He suggests that patent courts continue to use such equitable considerations in their decisionmaking and inform these traditional factors with precedent from copyright fair use cases.

Professor Donald Chisum goes a bit further. While arguing that there is no statutory basis for excluding algorithms from patentable subject matter, he also recognizes that patenting such fundamental building blocks may impede further progress if no use is privileged. For example, he finds it troubling that if a programmer were to experiment with a patented algorithm during research leading to the marketing of a non-infringing product, that programmer would be guilty of patent infringement. While he regrets the lack of a fair use provision in patent law, he stops short of proposing such an addition, perhaps in part because it was incidental to his main thesis on the patentability of algorithms.

enforcement of the patent when the improved technology is ready for commercial exploitation. On the other hand, if the subsequent researcher uses the patented invention to invent around the patent, developing a new technology that may be exploited without infringing the patent claims, the patent holder will have no means of extracting payment from the researcher at the commercial exploitation stage. In these circumstances, denying the patent holder a damage remedy for the research use would undermine the value of the patent monopoly and lead to unjust enrichment of the researcher.

94. See Grossman, supra note 15, at 245–46, 268–64 ("As relief flows to the patentee, and the more liberal attitude in the Patent and Trademark Office for qualifying subject matter [continues] ... courts may soon be faced with a request for relief in what appears to be a form of infringement that has practically no effect on the incentive purpose of the patent laws to promote investment based risk, and in fact, contributed to the progress of science and the useful arts or some form of technological innovation.").
95. See id. at 254.
96. See id. at 263.
98. See id. at 1017–18.
99. See id. at 1018–19.
2. The International View. — As domestic scholars continued to consider these topics, the United States negotiated the Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement as part of the General Agreement on Tariffs and Trade (GATT). While the United States pursued a relatively high protectionism agenda, the TRIPS Agreement recognizes the “principle that patents should not impair the advance of technology, and, consequently, that the effects of patents should not cover certain acts.”

While some European countries advocated enumeration of specific privileged uses in the TRIPS Agreement, the eventual compromise was reflected in the vaguer Article 30:

Members may provide limited rights to the exclusive rights conferred by a patent, provided that such exceptions do not unreasonably conflict with a normal exploitation of the patent and do not unreasonably prejudice the legitimate interests of the patent owner, taking account of the legitimate interests of third parties.

The meaning of this provision is unclear, and one writer describes it as “extremely manipulable.” When interpreted against the backdrop of the legislative debate and other proposals, it seems likely that it was intended to allow exceptions for infringements conducted for: (i) private, non-commercial purposes; (ii) research; (iii) experimentation for testing or improvement; and (iv) educational purposes.

The absence of a well-defined research and experimentation exemption in patent law is troublesome and especially so in the area of computer software development in which ‘playing around’ with prior efforts appears to be an important part of the research process. However, computer technology is not the only area in which research and experimental use of technology belonging to others is important. For example, such use would appear to be equally important to research in chemistry and biotechnology. The patent system as a whole is not perfectly clear, consistent, and logical, but the same can be said of other areas of the law, such as the tort system and the tax system.

Id. at 1019. This suggests that Professor Chism may find a fair use doctrine in patent law useful, at least in some circumstances.


102. See id. at 202–03 (describing German law and suggestions by European Community countries that particular exceptions be written into the agreement).


104. Straus, supra note 101, at 203.

105. See Carlos M. Correa, Patent Rights, in Intellectual Property and International Trade: The TRIPS Agreement 189, 208 (Carlos M. Correa & Abdulqawi A. Yusuf eds., 1998) (interpreting the provision “[b]ased on comparative law and other proposals,” including the WIPO draft Treaty Supplementing the Paris Convention, and stating other contexts in which infringement would likely be excused); see also Straus, supra note 101, at
Further, Article 8 permits member countries to "adopt measures necessary to protect public health and nutrition, and to promote the public interest in sectors of vital importance to their socio-economic and technological development . . . [and] measures . . . to prevent the abuse of intellectual property rights by right holders." 106 Any measures enacted pursuant to Article 8 must be "consistent with the provisions" of the TRIPS Agreement. 107

This states a strong international norm in favor of allowing socially beneficial infringements to occur. Moreover, the language of Article 30 is quite similar to that of Article 13 addressing exceptions to copyright rights. Article 13 states, "Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder." 108 The United States contends that its copyright fair use doctrine is permissible under Article 13, although some other countries disagree. 109 To the extent that Article 30 parallels Article 13, this suggests that some type of patent fair use is not only permissible but also expected under TRIPS. 110

202-03 (stating that the interpretation of Article 30 should be conducted against the backdrop of already existing exceptions, "particularly in those Member countries that played instrumental roles during the negotiations").

106. TRIPS Agreement, supra note 103, art. 8

107. Id.

108. Id. at art. 13.

109. See Tyler G. Newby, What's Fair Here Is Not Fair Everywhere: Does the American Fair Use Doctrine Violate International Copyright Law?, 51 Stan. L. Rev. 1693, 1648-62 (1999) (arguing that American fair use is permissible under Article 13 and describing U.S. responses to European Community and Australian inquiries as to how that could be the case). Admittedly, other countries would find it considerably ironic if the U.S. were to adopt a fair use principle in patent law. The U.S. has been particularly vocal in pressing for strong intellectual property rights and limiting, for example, reverse engineering exceptions to only those intended to achieve compatibility. See Crystal D. Talley, Note: Japan's Retreat from Reverse Engineering: An Unnecessary Surrender, 29 Cornell Int'l L.J. 807-809 (1996) (describing how, when Japan considered relaxing its copyright law to permit decompilation, the United States "denounced the Japanese proposal, proclaiming that it was 'contrary to international norms' and would 'set a dangerous precedent.'); see also, Pamela Samuelson, The U.S. Digital Agenda at WIPO, 37 Y. Int'l L. 669 (1997).

110. See generally Michael Blakeney, Trade-Related Aspects of Intellectual Property Rights: A Concise Guide to the TRIPS Agreement 87 (1996) (arguing that compulsory licensing in the public interest and use or manufacture for research and experimentation are permitted under Article 30); Michael Halewood, Regulating Patent Holders: Local Working Requirements and Compulsory Licenses at International Law, 35 Osgoode Hall L.J. 243 (arguing that nations retain a wide range of policy choices allowing them to limit a patentee's rights while still complying with TRIPS). But one commentator described a proposal that TRIPS negotiators adapt fair use for patent law as "rejected by both patent specialists and copyright specialists as a perversion of policy purposes in both their houses." Michael P. Ryan, The Function-Specific and Linkage-Bargain Diplomacy of International Intellectual Property Lawmaking, 19 U. Pa. J. Int'l Econ. L. 535, 565 (1998). However, the source that Ryan cites does not, in fact, state this proposition.
The main objective of the TRIPS Agreement was to compel developing countries to adopt minimum levels of intellectual property protection. However, some less developed countries (LDCs) that do not currently have industries that rely on intellectual property protection fear that even this “minimal” level will leave them at a disadvantage.\textsuperscript{111} Exports from the developed world that are protected by intellectual property rights may both exacerbate the LDCs’ trade deficits and become an obstacle to their developing domestic industry.\textsuperscript{112} The LDCs are therefore seeking to implement the TRIPS Agreement in the least restrictive way possible. In fact, international institutions are already drawing on this Article in suggesting that LDCs consider adopting fair use in patent law as consistent with their obligations under the TRIPS Agreement and preferable to more invasive alternatives like compulsory licenses.\textsuperscript{113}

3. Implementing fair use in patent law. — What should a patent fair use doctrine look like? Analyzing both existing law and the shared scholarly thought discussed above identifies gaps in current doctrine that may be profitably addressed in a fair use test. The five-part test proposed here emphasizes factors relevant to patent law and departs from the copyright model by authorizing courts to impose a fee on the fair user. This particular innovation helps to ensure that the patentee’s incentives are adequately protected.

a. A critical examination of existing doctrine. — Under current law, a court may excuse infringement under the reverse doctrine of equivalents or the experimental use exception. Excusing infringement under the former is justified because the infringer is a radical improver, the probability of bargaining breakdown is high, and society would bear large costs if the new invention were kept from it.\textsuperscript{114} This suggests that a comprehensive fair use doctrine should consider the nature of the advance that the infringing work represents, with major advances weighing in favor of the infringer.

Excusing infringement under the experimental use exception is highly unlikely to impact the patentee’s incentives because the infringer’s purpose is definitionally to satisfy his or her own “amusement.”\textsuperscript{115} This suggests that a fair use defense should consider whether the purpose of the infringement is commercial or not.

It also highlights a potential gap in patent law. What about cases in which the infringement is only indirectly commercial because it is con-

\textsuperscript{112} See id.
\textsuperscript{113} See e-mail from Ruth Okediji, Apr. 12, 2000 (on file with the Columbia Law Review).
\textsuperscript{114} See supra notes 63–66 and accompanying text (discussing the reverse doctrine of equivalents).
\textsuperscript{115} See supra notes 69–70 and accompanying text (discussing the experimental use exception).
ducted during pure research or as a step on the way to producing a non-infringing end product?

In the former case, the domestic and international thought discussed above argues that such infringement should be excused as fair. This research, if enjoined, would frustrate further progress without a benefit that should be legally cognizable accruing to the patentee. Fair use would help to overcome market defects attributable to high transaction costs, anti-patent motives, or externalities that prevent licensing in this context from occurring.

Harder questions arise when the research is directed toward the commercial end of marketing a non-infringing product. Professor Chisum and the international view seem to lean toward allowing even this infringement to proceed for free. Professor Eisenberg looks at things differently. With an eye toward maintaining the original patentee's incentives, she suggests that the research infringement be excused but the infringer required to pay the patentee. This is reminiscent of the Supreme Court's suggestion in *Auff-Rose* in the context of copyright fair use.\(^{116}\) Further, she argues that if the research results merely in an improvement rather than a non-infringing end-product, the infringer should seek a license under the blocking patents doctrine before it markets its product. She does not specifically answer the question of what happens to the improver if the original patentee refuses to grant a license, but implies that the improver's inability to market its product is the price of maintaining the patentee's incentives.

But the substantial improver who obtains a blocking patent may face the same bargaining breakdown that the radical improver does.\(^{117}\) However, no existing doctrine authorizes excusing the infringement if the improver markets its invention. While this result may generally be justified to maintain the original patentee's incentives, there may be cases where society would benefit from allowing the infringement to continue. A fair use defense should consider whether market defects exist that threaten the utility of the blocking patents doctrine in facilitating the licensing of improvements. It may then balance the impact on the patentee's incentives if the infringement were excused against the increase in social welfare if the improvement were made available.

Taken together, these strands of thought demonstrate that patent law is not adequately equipped to excuse infringement conducted with a direct or indirect commercial purpose. The Chisum view does not consider the original patentee's incentives while the Eisenberg approach does not address circumstances where market defects cause the blocking patents doctrine to fail in effecting a cross-license. The conflict between the two might best be resolved with a patent fair use defense that, more clearly than its copyright counterpart, bifurcates the questions of whether

---

116. See supra note 56.
117. See supra notes 67-68 (discussing the blocking patents doctrine).
the infringement should be excused and whether the infringer should pay for the use.

Professor Chisum's concerns regarding the patentability of algorithms raise another point. All patented inventions are not created equal. Follow-on innovators have more need to make an infringing use of some inventions than others. A patent fair use defense should therefore also consider the nature of the patented work.

Finally, none of the authorities supports excusing commercial infringement that occurs in the marketing of a directly infringing product. But copyright fair use warns that there may be cases where such infringement is appropriately excused. The lack of any universal principle accounting for differences between the copyright and patent incentive schemes suggests that there may be such cases in patent law as well. The fair use defense should be flexible enough to leave a court with room to excuse even direct commercial infringement when appropriate.

b. The test for excusing infringement as fair. — Practically, courts will need defined factors to guide them in applying a patent fair use defense. The current lack of a coherent theoretical framework may help to account for their reluctance to employ the tools that they already have (particularly the experimental use doctrine and reverse doctrine of equivalents) to excuse some infringements. In the absence of a defined test, they are equally unlikely to employ a fair use defense that encompasses those tools and offers even a broader haven for the infringer.

The preceding analysis identified five factors relevant to a fair use finding: (i) the nature of the advance represented by the infringement; (ii) the purpose of the infringing use; (iii) the nature and strength of the market failure that prevents a license from being concluded; (iv) the impact of the use on the patentee's incentives and overall social welfare; and (v) the nature of the patented work. While this test resembles that of copyright fair use, it diverges to reflect the different incentive scheme of patent. However, copyright fair use theory and precedent is still useful in understanding how to implement the test.

118. The Copyright Act sets forth four factors for courts to use in evaluating a claim of fair use: "(1) the purpose and character of the use . . . (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work." 17 U.S.C. § 107 (1994 & Supp. III 1997). As detailed in the text, the patent fair use test contains colorable approximations of factors (1), (2), and (4). The third copyright factor does not play a role in patent fair use because a patent infringer must make, use, or sell the patented invention defined by the claims in its entirety. In contrast, in copyright, a person may infringe by taking only part of the copyrighted work. There is a doctrine of de minimis infringement under patent law which may support excusing infringement under certain circumstances or refusing to grant a remedy where infringement is found. See Grossman, supra note 15, at 249–50 (describing the doctrine). This concept is taken into account under the proposed test by requiring the courts to consider the purpose of the use, which indirectly would consider whether the infringement were de minimis or not.
i. Factor One: The nature of the advance represented by the infringing work.
— Under this factor, the court would consider whether the infringing invention represents a minor or major step forward. The more significant the advance, the more likely the public welfare would be substantially advanced if it were made available. Thus, a major technological leap forward should weigh in favor of fair use while a minor advance should weigh against it.

That an infringement does not result in a major improvement would not bar a fair use finding. This factor is most helpful when the potential fair user markets a directly infringing product. It is less probative where the infringement is conducted during research or as a step in marketing a non-infringing product. It has its least force in the case of network markets where compatibility with the dominant standard is required for market entry. There, it often is not a viable strategy to improve the patented invention because changes to it would defeat the compatibility required to compete effectively.119

ii. Factor Two: The purpose of the infringing use. — Here a court would assess whether or not the infringing use is non-commercial or directly or indirectly commercial. As in copyright law, a non-commercial use should weigh in favor of fair use while a commercial use weighs against it. The commercial use is much more likely to harm the patentee’s incentives without a corresponding increase in social welfare.

However, patent courts should make the same distinction as their copyright counterparts. Copyright courts are more hospitable to infringement that is conducted as an intermediate step in producing a non-infringing end product than towards a directly infringing product.120 This reflects their attempts to calibrate the scope of the right appropriately to encourage follow-on uses that do not usurp the market in the same way as a directly infringing end product does.

iii. Factor Three: The nature and strength of the market failure that frustrates licensing. — A court should consider what type(s) of market failure is/are present that prevents a license from being concluded in the particular case. While this is not explicitly one of the copyright fair use factors, it has implicitly guided at least some part of copyright fair use jurisprudence. Making it explicit in the area of patent may help courts better integrate theory into decisionmaking.121 Courts should identify the relevant market defect whether it be high transaction costs, externalities, or

119. See generally infra Part II.A (describing network markets and why the reverse doctrine of equivalents will not enable competition).
120. See infra Part II.A.2 (discussing the copyright video game cases).
121. Congress is considering adopting an analogous approach in one version of database protection legislation. See H.R. 1858, 106th Cong. § 106 (2) (providing that one of the factors to be weighed in considering whether an entity has misused its database protection is “the extent to which information contained in a database is the sole source of the information contained therein is made available through licensing or sale on reasonable terms and conditions”). In contrast, section 271(d) of the Patent Act provides that it is not misuse to refuse to license unilaterally. 35 U.S.C. § 271(d)(4) (1994).
something else, and explain how it impacts the market, particularly for innovation.

At this point, it is appropriate to recall the earlier warning—that refusals to license in copyright law should generally be respected unless the copyright owner’s refusal is unrelated to realizing copyright’s incentive goals.122 The same rationale holds for patent. Courts should look for evidence of market defects that strongly indicate that the failure to license in the particular case is inefficient or motivated by an anti-patent intent.

iv. Factor Four: The impact of the use on incentives and social welfare. — In applying a similar factor under copyright law,123 courts focus on whether, if the infringer’s use were to become widespread, it would adversely impact the market for the copyrighted work.124 This helps to explain why courts are more hospitable to copyright fair use claims when the infringer has “transformed” the work in some way.125 Their ultimate concern is how to calibrate the balance between the copyright owner’s rights and those of the public against the backdrop of the statutory goals. As the Second Circuit has said:

[W]here a claim of fair use is made, a balance must sometimes be struck between the benefit the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied. The less adverse effect that an alleged infringing use has on the copyright owner’s expectation of gain, the less public benefit need be shown to justify the use.126

Likewise, a court assessing a patent fair use claim must focus on the likely effect of its holding on social welfare. In turn, this implies that courts should assess the social benefit to be gained by allowing the infringement balanced against the costs to the patentee, including the impact on incentives to invent. Harm to the patentee is likely to be greater

Introducing this consideration into fair use is appropriate because it is relevant to the question of social welfare and not otherwise available for examination under the statute.

122. See supra text accompanying note 51.

123. See 17 U.S.C. § 107(4) (1994 & Supp. III 1997) (listing “the effect of the use upon the potential market for or value of the copyrighted work” as a factor a court may consider in addressing a fair use claim).

124. See Campbell v. Acuff-Rose Music, 510 U.S. 569, 590 (1994) (noting that the fourth factor “requires courts to consider not only the extent of market harm caused by the particular actions of the alleged infringer, but also ‘whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market’ for the original . . . The enquiry must take account not only of the harm to the original but also of harm to the market for derivative works’”) (citations omitted); Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1523 (9th Cir. 1993) (“We must, of course, inquire whether, if [the challenged use] should become widespread, it would adversely affect the potential market for the copyrighted work . . . by diminishing potential sales, interfering with marketability, or usurping the market.”) (citing Sony Corp. v. Universal City Studios, 464 U.S. 417, 451 (1984)).

125. See Loren, supra note 15, at 30–31 (noting and criticizing courts’ focus on whether or not a use is transformative in conducting a fair use inquiry).

when the infringement leads to a competitive product and it will be greatest when that product is also directly infringing.

The courts should focus on the nature of both R&D and product-market competition in the particular industry. For example, if the R&D investment giving rise to the patent is of a type that generates significant spillovers, making it difficult for the patentee to capture the value associated with its innovation, allowing the infringement to proceed (particularly for free) might unduly reduce innovation incentives. Similarly, if the R&D investment is quite large in absolute dollars, the patentee may require a certain lead time in the market to allow it to recoup both its investment and a reasonable return thereon.127

This is the most important fair use factor because it focuses on ensuring that the goals of the patent system are adequately protected. A court must integrate its findings regarding market defects into its incentives analysis under this factor. If there are no significant market defects, then the court should be more willing to refuse to label a use "fair" even though the public might benefit from it.

v. Factor Five: The nature of the patented work. — As a copyright court considers the nature of the copyrighted work in a fair use inquiry, so too should a patent court examine closely the subject matter of the patent. In copyright law, the scope of the exclusive right is "thinner" for fact-based or utilitarian works than for highly creative works.128 This translates into broader fair use rights for works with thin copyrights. In the case of patent, while utility, novelty, and non-obviousness are always conditions of protection, an invention may, in fact, be more or less pioneering. Any fair use right should be broader where the infringed invention is a relatively small advance over the prior art and narrower where it is a major step forward. In the former case, the likelihood of a patentee's acting out of an anti-patent motive is much greater.129

The irony is that there may be a direct relationship between the degree of inventiveness of the first innovation and the need for follow-on inventors to infringe the patent in adding to the store of knowledge. But if fair use is construed narrowly for pioneering inventions, the follow-on inventor will not be able to infringe. It must either obtain a license or wait until the patent expires.

This may simply be the price for maintaining the patentee's incentives. However, there may be some cases when the overall test weighs in favor of the infringer despite the pioneering status of the original invention. Generally, though, fair use is more likely to protect the improver who adds a meaningful contribution to something less than a pioneering

127. See generally Reichman, supra note 34, at 2516 (noting the importance of lead time in recouping investment).


129. See infra note 180 and accompanying text (discussing "disproportionate leverage").
invention. This does not mean that fair use will be an empty right. Many
inventions are successful not because they are so ingenious but rather
because they enter the market at a particularly propitious time.

The five-factor test authorizes a court to balance considerations rele-
vant to patent law in deciding whether to excuse a use as fair. The most
important factors are the third and fourth which emphasize the reality of
market conditions and the impact on the intellectual property balance.
As in copyright, no one factor would be determinative and fair use would
be an equitable and affirmative defense with the burden of proof on the
infringer.

c. Fair use: royalty-free or royalty-bearing? — If the court decides that
the infringement is “fair,” it then must consider whether or not the in-
fringer should compensate the patentee.\textsuperscript{130} The court can use the five-
factor test to help in this inquiry, focusing primarily on the market failure
in the particular case that justifies a finding of fair use. In at least some
cases, that rationale will not support a royalty-free license.

In the copyright cases, the defect is often one of high transaction
costs that prevent private bargaining. Those same costs may also make it
prohibitively expensive to set up a mechanism for a fair user to pay a fee.
In patent cases, particularly when only one patentee is involved, it seems
less likely that excusing non-payment because of the high transaction
costs in setting up a payment system would be implicated.

Moreover, copyright sometimes excuses payment because the fair
use is of a type, like criticism or responding to it, that policymakers be-
lieve should not be commodified. The same concern is simply unlikely to
arise with respect to patented inventions and therefore would not justify a
“free” fair use.

However, copyright also excuses payment when the infringing use is
socially beneficial and the infringer simply cannot pay the price, usually
because it cannot capture the externalities associated with its use. Cer-
tainly, educational and other non-profit uses like pure research often fall
into this category. Alternatively, no charge may make sense if there is no

\textsuperscript{130} There is some caselaw authority for such an approach. For example, in the City
of Milwaukee v. Activated Sludge, Inc., 69 F.2d 557 (7th Cir. 1994), the Seventh Circuit
refused to enjoin the city from infringing a patent related to sewage treatment:

If... the injunction ordered by the trial court is made permanent in this case, it
would close the sewage plant, leaving the entire community without any means
for the disposal of raw sewage other than running it into Lake Michigan, thereby
polluting its waters and endangering the health and lives of that and other
adjoining communities... [W]here, as here, the health and the lives of more
than half a million people are involved, we think no risk should be taken.
69 F.2d at 593. While this is an extreme example in which the public benefit from
infringing the patent was quite large, it illustrates that fair use does not have to be the
equivalent of a royalty free compulsory license. A court might label a use “fair” but require
the infringer to pay a fee to safeguard the patentee’s incentives.
harm to the patentee.\textsuperscript{131} Moreover, the patentee should not be compensated for behavior that seeks to retard the very purpose of the patent laws.

\textit{d. The implementation details—Congress or the courts?} — An important question is whether Congress or the judiciary is the best institution to adopt the five-factor test. Because Congress possesses the resources to make detailed empirical studies, can provide courts with guidance on interpretation through an extensive legislative history, and clearly has the power to adopt fair use under the Constitution’s copyright and patent clause, it is the preferable implementing institution.

The question is whether it has the political will to weaken the exclusive patent rights. If it does not, courts may be called on to do so as a matter of common law. Historically, courts in intellectual property cases have adjusted the law to address changing conditions. For example, copyright’s fair use doctrine originated with the judiciary over 150 years before its codification in the Copyright Act.\textsuperscript{132} Likewise, the important threshold requirement of nonobviousness in patent law arose at common law prior to its implementation in section 103 of the Patent Act.\textsuperscript{133} This tradition suggests that courts have the power to adopt a fair use doctrine although they, like Congress, may lack the will to do so.

However, recent Supreme Court precedent casts doubt on whether, in fact, courts retain this traditional latitude in formulating intellectual property doctrine. Under recent decisions, it is unclear whether the judiciary has the power to adopt a new rule of decision when there is no specific statutory term inviting such interpretation.\textsuperscript{134} While one might argue that fair use reflects concepts that have always been implicit in patent law, it strains credulity to suggest that there is an express term in the Patent Act inviting interpretation with the detailed test this Article proposes.

Regardless of what institution undertakes the job of reforming the law and in what manner, the point remains. Patents that are overbroad

\textsuperscript{131} See generally Grossman, supra note 15, at 263 (“[A]ny enforcement of a patent should depend on whether or not the rights granted to the patentee have been adversely affected.”).

\textsuperscript{132} See supra note 38.

\textsuperscript{133} See Chism et al., supra note 8 at 531 (tracing non-obviousness to an 1850 judicial opinion).

\textsuperscript{134} See, e.g., Burlington Industries, Inc. v. Ellerth, 524 U.S. 742, 755 (1998) (emphasizing that the Court, in defining “agent” under Title VII of the Civil Rights Act, was engaged in “statutory interpretation pursuant to congressional direction” rather than “judicial ‘creation’ of a special federal rule of decision”). On the other hand, Atherton v. FDIC, 519 U.S. 213 (1997), suggests that this judicial creation of a federal rule of decision is primarily a concern in cases where the federal rule would displace state law. See 519 U.S. at 217-26 (holding that there is no general federal common law defining the standard of care for officers and directors of federally insured savings institutions). Atherton suggests that because there is no state patent system providing a rule of decision, courts may retain power to implement fair use on their own. However, \textit{Burlington} may mean that courts cannot invent new doctrines even in interpreting federal statutes unless there is an express term in the statute inviting such interpretation.
either when initially issued or that become so over time as market conditions change, impose costs on society. The number of such patents is likely increasing. The patent fair use of this Article is a bold proposal that may stand little chance of enactment either by Congress or the courts (if they even have the power to do so).\textsuperscript{135} Regardless, it performs a useful service by identifying the shortcomings of existing doctrine and is at least a starting point for discussion if not a blueprint for legislative change.

II. EXPANDING THE THEORY AND INTEGRATING IT INTO PRACTICE: APPLICATIONS OF FAIR USE IN PATENT LAW

The desirability of incorporating a fair use doctrine into patent law depends not merely on its having sound theoretical underpinnings, but also on an ability to implement it in a cost-effective manner. The following draws on economic theory to explain why market failures that frustrate licensing are increasing. It then applies this insight to different contexts, giving examples of instances in which fair use could perform a valuable function.

A. Patent Fair Use for Software Application Programming Interfaces

Examining the functioning of patent law in the software industry is not only a timely exercise but also a peculiarly appropriate one. The connectivity components of operating systems that specify how a particular operating system and its applications communicate are an interesting example of subject matter that has migrated from one incentive scheme to another. Over the years, copyright protection for these components (called application programming interfaces (APIs))\textsuperscript{136} has receded in favor of patent law. An understanding of the software market informs the narrative of copyright courts’ eventual movement to hold APIs protectable, if at all, by patent law. But their relegation of APIs to the entirely different incentive structure of patent law left some questions unanswered. The patent right, while more difficult to obtain, is less amenable to temporal adjustment than a copyright.

The desirability of fine-tuning rights in APIs as the market develops stems in part from the particular characteristics of the market for operating systems software. The following briefly explains why the network features of the software market may make a case for weakening intellectual property protection for APIs. It then examines the lessons that the copyright treatment of APIs holds for patent law, particularly in deploying its

\textsuperscript{135} See generally Merges, Impossible Patents, supra note 3, at 588 (proposing reforms to the patent process and noting that while there may be little chance that the relevant institutions will act, it is still worthwhile to bring both problems and solutions to policymakers’ attentions).

\textsuperscript{136} See Microsoft Findings of Fact, No. 98-1233 (TFP), ¶ 2 (D.D.C. filed Nov. 5, 1999) [hereinafter Microsoft Findings of Fact] (defining an API as “synapses at which the developer of an application can connect to invoke pre-fabricated blocks of code in the operating system”).
new fair use defense. It argues that in some cases, courts should employ fair use to excuse infringements conducted during reverse engineering to produce software compatible with the dominant operating system.

1. Network Effects and the Case for Weaker Intellectual Property Rights. — The economics of network markets have greatly influenced judicial opinions, both implicitly and explicitly, as protection for APIs has moved from copyright law to patent law. Generally, network effects are present when a consumer’s utility associated with a good increases as others also purchase it. In an “actual” network like the telephone system, the network effect is direct as the product’s entire value inheres in enabling communication among product owners. In contrast, “[g]oods constitute virtual networks when they provide inherent value to consumers that increases with the number of additional users of identical and/or interoperable goods.” For example, in part because applications may generally run only on the operating system for which they are designed, operating systems are characterized by network effects—as the number of applications running on an operating system increases, the value of that system also goes up, causing more developers to write applications for it and more consumers to adopt it, and so on. This phenomenon is called a “positive feedback effect,” and it helps to explain why network markets are characterized by large first-mover advantages.

137. See Michael L. Katz & Carl Shapiro, Network Externalities, Competition, and Compatibility, 75 Amer. Econ. Rev. 424, 424 (1985) [hereinafter Katz & Shapiro, Network Externalities]; see also Microsoft Findings of Fact, supra note 136, at ¶ 39 (stating that “[a] positive network effect is a phenomenon by which the attractiveness of a product increases with the number of people using it” and citing the Microsoft Windows operating system as an example of such a product).

138. See Mark A. Lemley & David McGowan, Legal Implications of Network Economic Effects, 86 Cal. L. Rev. 479, 488–89 (1998) (“The value of the telephone or fax machine is already purchased increases with each additional purchaser, so long as all machines operate on the same standards and the network infrastructure is capable of processing all member communications reliably.”). Under the textual definition, actual networks are limited to communications markets. See id. at 489.

139. Id. at 491; see also Peter S. Menell, An Epitaph for Traditional Copyright Protection of Network Features of Computer Software, 43 Antitrust Bull. 651, 657 (1998) (“The defining feature of virtual networks is that the value of a product depends significantly on the value of the system in which it is a part. A critical determinant of the value of a product, therefore, is the range and value of components with which it is interoperable.”).

140. See Microsoft Findings of Fact, supra note 136, at ¶ 4 (noting that “an application written for one operating system will not function on another” unless it is first adapted, or “ported,” to the APIs of the other operating system”).

141. Network effects, although somewhat less strong than in the case of operating systems, also characterize the market for application programs. As the Lotus case discussed below illustrates in the electronic spreadsheet context, see infra note 195, when a user invests time and effort in learning how to use a particular program, he or she is unlikely to adopt a new, improved product that requires different skills and/or that cannot process the user's files.

142. See Microsoft Findings of Fact, supra note 136, at ¶ 30–40 (explaining how this effect has made the number of applications written for Microsoft Windows a substantial
The positive feedback effect characteristic of both actual and virtual networks causes the market to tip to the adoption of one dominant standard for connectivity. For example, Microsoft’s Windows’ API is the dominant standard in the PC operating systems market. Compatibility with the standard is the key to a new market entrant’s ability to compete. No PC operating system would be successful if it could not run the same set of applications that Microsoft’s Windows does. The operating system, then, must be “horizontally compatible” with Windows, implementing all of its APIs and potentially usurping Microsoft’s customers by offering them an alternative. Likewise, for an application to be a viable competitor in the PC market, it must be “vertically compatible” with Windows by running successfully on it. As more applications write to the Windows API, the demand for the underlying operating system implementing those APIs, whether furnished by Microsoft or a horizontally compatible competitor, increases.143

Policymakers must decide what legal rules are most likely to ensure optimal levels of R&D and product-market competition. They have essentially three choices: (i) an incompatibility regime in which strong intellectual property rights enable a single dominant firm to control the standard and who implements it; (ii) a compatibility system in which relatively weak intellectual property rights allow the marketplace to implement the standard; or (iii) some combination thereof. A brief analysis of the two extremes reveals that this last option makes the most sense.

A market under an incompatibility regime is likely to be characterized by strong competition to set the industry standard as firms fight to win the entire market: The prospect of this large reward may induce socially excessive investments in R&D.144 Once the standard is set, R&D competition will decline as firms are weeded out and others refuse to

__________

barrier to entry into the operating systems market and stating that “[w]hat for Microsoft is a positive feedback loop is for would-be competitors a vicious cycle”; Joseph Farrell & Michael L. Katz, The Effects of Antitrust and Intellectual Property Law on Compatibility and Innovation, 45 Antitrust Bull. 609, 610 (1998); Menell, supra note 159, at 658; see also Joshua Quittner, Fringe Benefits, Time, Nov. 22, 1999, at 112 (noting that about 70,000 applications are available for Windows compared to 1,000 for BeOS, 2,500 for IBM’s OS/2, and 12,000 for Apple Computer’s Mac OS and questioning whether Linux, the free operating system circulated on the Internet and distributed by some computer manufacturers, can present a viable alternative to Windows). For an early case in which a competitor sought compatibility with an operating system running a large number of applications, see Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240 (3d Cir. 1988), (holding the defendant Franklin liable for infringing Apple’s copyrights by copying Apple’s operating system in an attempt to take advantage of the number of applications running on that system).

143. See Lemley & McGowan, supra note 138, at 524 (defining horizontal and vertical compatibility).

144. See Farrell & Katz, supra note 142, at 638 (noting that if consumer expectations follow quality and compatibility offers efficient incentives to innovate, excessive innovation can occur under an incompatibility regime). However, the high risk of network markets may mitigate this effect. See infra note 161.
enter because their goods would be incompatible with the dominant firm’s.\textsuperscript{145} While even the standard-setting monopolist will innovate, it may not be as responsive to incentives as firms in competition would be, slowing the pace of innovation to guard against the erosion of existing profit streams.\textsuperscript{146} Competition focuses on installed base and size of the network because consumers’ purchasing decisions are influenced by their expectations as to what product will win the standards competition. They fear being stranded on the losing, incompatible system and incurring the costs of switching to the winner.\textsuperscript{147}

Under a compatibility regime, because the standard (and therefore the network) is open to all firms, competition centers around traditional indicators like price and quality rather than installed base and expected size of the network.\textsuperscript{148} At the outset, competition is less fierce because tipping will not result in one firm’s obtaining the market; however, competition intensifies later, once the standard is set.\textsuperscript{149} This latter period may be quite long because only a radical innovation will unseat a durable standard.\textsuperscript{150} Under a compatibility regime, therefore, firms face lower

\textsuperscript{145} See id. at 638–40. But see Michael L. Katz & Karl Shapiro, Systems Competition and Network Effects, 8 J. Econ. Persp. 93, 108 (1994) (hereinafter Katz and Shapiro, Systems Competition) (noting that there are circumstances in which network markets might be biased toward newer, incompatible standards because buyers of the incompatible product ignore the standing costs they impose on prior buyers of the original standard).

\textsuperscript{146} See Farrell & Katz, supra note 142, at 639, 647; see also Menell, supra note 139, at 674 (noting that “monopolists often have less incentive to develop drastic improvements in their technology because such innovations displace the monopolist’s existing profits”).

\textsuperscript{147} See Katz & Shapiro, Systems Competition, supra note 145, at 94 (“Once a certain system is chosen, switching suppliers is costly because new relation-specific investments have to be made. In such a situation, systems that are expected to be popular—and thus have widely available components—will be more popular for that very reason.”); see also Microsoft Findings of Fact, supra note 136, at ¶ 28 (noting that the costs of switching from an Intel PC operating system to one that is not compatible with Intel are quite high since users would need to learn the new system, purchase new applications, and replace files stored under the old system).

\textsuperscript{148} See Farrell & Katz, supra note 142, at 611.

\textsuperscript{149} See Katz & Shapiro, Systems Competition, supra note 145, at 111. A market in which the standard is open is less risky. Innovation incentives to set the standard may not be sub-optimal because this lower risk positively affects the expected value of the investment.

\textsuperscript{150} See Microsoft Findings of Fact, supra note 136, at ¶ 18, 27, 33, 56 (emphasizing that alternatives to programming for Microsoft’s Windows APIs are unlikely to be realistic options for application developers in the next few years); see also Microsoft Corp., 87 F. Supp. 2d at 35 (citing the Findings of Fact indicating that no competitor could enter the
barriers to entry than under an incompatibility system where the inability to work with the dominant system can be a large impediment to market entry.\textsuperscript{151}

Because network markets are heavily path-dependent, there is some risk that the "winning" standard will be both inferior and durable, locking consumers into a sub-optimal system regardless of the legal regime.\textsuperscript{152} However, this risk may be somewhat lower under a compatibility system because competitors often join to set a technologically superior standard and compete on quality and price in offering products implementing and conforming to that standard.

For this reason, and because innovation competition is enhanced after the standard is set, many commentators assume that compatibility, with its relatively weak intellectual property protection, is efficient and, accordingly, legal rules should facilitate it.\textsuperscript{153} Unfortunately, although the existence of networks is nothing new, economists have done little work to support this assumption. From an intellectual property perspective, this is particularly troubling because "the greatest difference between systems markets and other markets may arise because firms' innovation incentives are altered by network considerations."\textsuperscript{154}

In particular, network markets may be a notable exception to the general rule that the grant of exclusive rights under either the copyright or patent law does not necessarily—or even usually—result in the relevant market as a viable alternative to Microsoft "within a reasonably short period of time".

\textsuperscript{151} See Microsoft Corp., 87 F. Supp. 2d at 35 (noting that "the application barrier to entry protects Microsoft's dominant market share").

\textsuperscript{152} The classic example is, of course, the QWERTY keyboard. See Paul A. David, Clio and the Economics of QWERTY, 75 Am. Econ. Rev. No. 2, 332, 332-33 (1985). But see S. J. Liebwater & Stephen E. Margolis, Should Technology Choice Be A Concern for Antitrust Policy?, 9 Harv. J.L. & Tech. 283 (1996) (refuting the claim that the QWERTY keyboard is an example of path dependence by identifying errors in the factual account of how the keyboard became dominant). See generally Mark J. Roe, Chaos and Evolution in Law and Economics, 109 Harv. L. Rev. 641, 647 (1996) (noting that it is unclear whether path dependence often leads to "highly inefficient structures that society cannot eliminate," and stating that even if it does, "we cannot do much about it"). Cf. Microsoft Corp., 87 F. Supp. 2d at 40 (rejecting Microsoft's contention that it integrated its Internet browser (Internet Explorer) into the Windows operating system with the intent to benefit consumers and noting that Internet Explorer is "not demonstrably the current best of breed Web browser, nor is it likely to be at any time in the immediate future"). However, not all network markets will tip to the adoption of a single standard: If competing systems emphasize different product attributes on which varying sets of consumers place value, the market may support more than one network. See Katz & Shapiro, Systems Competition, supra note 145, at 106 ("Consumer heterogeneity and product differentiation tend to limit tipping and sustain multiple networks.").

\textsuperscript{153} See Farrell & Katz, supra note 142, at 641 ("Many people appear to believe intuitively that compatibility is more conducive to competition and thus public policy should promote or mandate compatibility.").

\textsuperscript{154} Katz & Shapiro, Systems Competition, supra note 145, at 106 (identifying "innovation in systems markets as a promising area for future research").
rightholder’s obtaining an economic monopoly in an antitrust sense. Moreover, while leveraging power from one market to another may be a dubious strategy in conventional markets, it may be a quite rational one in network markets—and it may be facilitated by the ownership of intellectual property rights.

The superficial case for weakening intellectual property rights in network industries focuses on the externalities that characterize network markets. An individual considers only its own costs and benefits in deciding whether or not to join the network. But a decision to join increases the network’s value to existing members (a positive externality) while a decision not to enter or to join another network decreases it (a negative externality). Intellectual property rights allow a rightholder to price above marginal cost. The negative externality resulting from a refusal to purchase because the price is too high makes that refusal more costly than in a conventional market, thus arguably making the case for weaker intellectual property protection in network markets.

Moreover, because significant first-mover advantages and returns larger than those of conventional markets often characterize network markets, investors may not need the inducement of intellectual property rights at all to encourage them to fund innovation. In other words, while information may suffer from a “public goods” problem, the particular

155. Antitrust law defines an economic monopoly as the ability to raise price above the competitive level or to exclude competitors. See United States v. E. I. Du Pont De Nemours & Co., 351 U.S. 377, 391 (1956). Often, substitutes exist for the copyrighted or patented product that prevent the rightholder from achieving such an economic monopoly. See Merges, Impossible Patents, supra note 282, at 594-35.

156. See Lemley & McGowan, supra note 138, at 496 & n.54 (noting that the conventional wisdom, although debatable, held that leveraging across markets would be unlikely because a certain amount of monopoly power “can extract only a given amount of revenue” despite the presence of two markets but that “a firm with a strong position in a network market might have enhanced ability—although perhaps less desire—to leverage into a non-network market”); see also Microsoft Corp., 87 F. Supp. 2d at 45-46 (discussing Microsoft’s alleged scheme to leverage its operating systems monopoly, obtained in part through Microsoft’s ownership of copyrights, into the Web browser market).

157. See Menell, supra note 139, at 656.

158. See, e.g., Lemley, Improvement, supra note 10, at 996 (noting that exclusive rights “prevent competition in the sale of the particular work or invention covered by the intellectual property right, and therefore allow the intellectual property owner to raise the price of that work above the marginal cost of reproducing it”).

159. See Joseph Farrell, Argument for Weaker Intellectual Property Protection in Network Industries, in Standards Policy for Information Infrastructure 368, 369 (Brian Kahin & Janet Abbate eds. 1995) (explaining also that in his static model that does not account for the potential benefits of penetration pricing, the right’s scope may expand beyond its temporal and literal bounds). For example, Bell was able to leverage a 17-year patent into long-term control over the telephone market. See id. at 370. “Network effects may expand the effective horizontal scope of a patent. For instance, if a patent or copyright explicitly covers only one particular way of doing things, network effects may make users reluctant to buy a different product, even if it is not inherently inferior. Thus, the effective scope might cover a whole market, even if the ostensible scope covers only a narrowly defined product.” Id.
characteristics of network markets may solve those problems, rendering intervention through the grant of exclusive rights not only unnecessary but also harmful, because of the costs those rights impose.\textsuperscript{160}

But because network markets are also riskier than non-network markets,\textsuperscript{161} investors may demand some measure of exclusivity in the form of intellectual property rights to be willing to take such a gamble. Additionally, to date, investors have funded innovation in network markets in anticipation that intellectual property protection will be available for at least those parts of their products that have traditionally been protected in non-network markets. There is simply no empirical basis from which to conclude that investors would have committed funds in the absence of such intellectual property rights.

While the arguments for “weaker” intellectual property rights in network markets fall short of supporting no rights at all, they are more convincing in contending that the rights do not need to be as strong as traditionally conceived. In addition to the arguments in favor of a compatibility regime set forth above, it is often the case that in a competition among several firms to set a standard that will be protected by some type of intellectual property right, the winner is awarded the entire market even if the victory is a narrow one: This narrow winner is likely to receive a reward that substantially exceeds the value of its incremental contribution to the standards competition.\textsuperscript{162}

Unfortunately, however, at the time that a rightholder seeks protection, the nature of the market and the strength of network effects, if any, are unlikely to be reliably known. A categorical rule denying protection and effectively imposing a compatibility regime may induce sub-optimal incentives for initial innovation. But a strict incompatibility regime might induce initial over-investment followed by double dominance in the R&D and product markets and decreased follow-on R&D investment once a standard is established.

\textsuperscript{160} See generally Lemley & McGowan, supra note 138, at 534 (“One might rebut the traditional case for intellectual property in network markets . . . if one can demonstrate that the network effects themselves will ensure an adequate return to the initial creator even absent intellectual property protection.”). The intellectual property system is anything but costless to maintain. There are public administrative costs in setting up and maintaining the system, as well as costs in enforcing the rights through the judicial system. There are private costs to the rightholders in obtaining, maintaining, and enforcing the right (particularly patent), and social costs associated with either over- or underprotection relative to the social optimum. See supra note 16.

\textsuperscript{161} See Menell, supra note 139, at 665. Because of this high risk, a firm with an established reputation and brand identity and superior technology is more likely than a less-established firm to favor an incompatibility regime. See Katz & Shapiro, Systems Competition, supra note 145, at 111 (noting that because of the large returns in a network market, a firm confident of success will prefer incompatibility to capture those returns and setting forth reasons why a firm might be so confident); Menell, supra note 139, at 665–67 (outlining strategies a firm might choose to pursue and considering the factors that will influence a firm’s choice).

\textsuperscript{162} See Farrell, supra note 159, at 371.
This suggests that the compatibility paradigm should be dynamic, beginning with an incompatibility regime, and changing to compatibility when more information about market conditions becomes available. For example, if consumer expectations regarding what is or will be the winning standard were to become stubborn, and the dominant firm were to engage in conduct directed toward controlling other product markets or the innovation market, courts could open the standard to competition. Any incentive for initial over-investment in R&D would be mitigated by the knowledge that the standard might eventually be opened to competition.

A dynamic compatibility regime may be best suited to defining the appropriate level of rights in APIs. The creation of some interfaces may entail large costs and an interface may be valuable compared to others in the market. However, many interfaces have a low intrinsic value but a high market price after consumers begin to adopt them. This suggests that "arbitrary or obvious interfaces, plausibly, should receive relatively little protection. . . . On the other hand, interfaces that reflect a higher level of achievement should receive more protection and authorities should be more reluctant to give competitors rights of access to the standards." Thus, rather than focusing on weakening protection directly, such a scheme would achieve the same effect by raising the threshold for obtaining exclusive rights.

The recommendation that only non-obvious interfaces receive protection conjures up the specter of patent law, with its more exacting threshold requirements (including non-obviousness), as the relevant intellectual property scheme rather than copyright. But a patented interface may become a standard over time as much as a copyrighted one, and patent law generally provides stronger rights and is less amenable to a temporal adjustment of those rights than copyright.

The implementation of a dynamic compatibility regime is less likely if APIs are protected by patent law, which lacks the fair use doctrine that copyright employs. While it may be appropriate that APIs meet the

---

163. Preliminary studies indicate that the crucial variable is likely consumer expectations. See generally Farrell & Katz, supra note 142, at 676–21 (explaining the importance of consumer expectations and modeling compatibility and incompatibility under different assumptions regarding what metrics, such as price or quality, primarily influence those expectations); see also generally Microsoft Findings of Fact, supra note 136, at ¶ 50 (noting the difficulty of another competitor's overcoming the "cycle of consumer preferences and developer incentives" that locks users into Microsoft's Windows operating system).

164. See Farrell & Katz, supra note 142, at 643.

165. See id.

166. Id. at 649. Farrell & Katz emphasize that there may be cases in which even an obvious interface should receive protection: "If genuinely innovative ideas developed by a dominant firm are difficult to protect directly against imitation, but can be protected indirectly through control of an interface, then opening up the interface might inadvertently reduce innovation incentives." Id. They also warn that their analysis is rough and it is therefore too early to derive firm policy principles from it. See id. at 650.
threshold requirements of patent law before being protected,\textsuperscript{167} it may also be that over time, society would benefit from some weakening of the patent right that existing scope limiting doctrines will not provide. The following addresses the relatively narrow question of API protection, briefly reviewing the development of rules favoring a compatibility regime in copyright law, and explaining how the proposed patent fair use defense could respond to the particular problems created by network markets.

2. The Copyright Example. — An appreciation of the consequences of network effects for copyright owners, their competitors, and consumers has led courts to use a variety of scope limiting doctrines, including fair use, to refine the nature of the exclusive rights. The copyright theory as it has developed in the courts illustrates how the economics of the software market is shaping the law, what questions the courts have left unanswered, and also informs the application of fair use in patent law.

The groundbreaking case addressing copyright protection for the functional components of software was \textit{Computer Associates International, Inc. v. Altai, Inc.}\textsuperscript{168} Before \textit{Computer Associates}, courts were in some disarray, with the primary precedent suggesting that APIs were copyrightable so long as they were not necessary to implement the program's overall function (or idea).\textsuperscript{169} This rule effectively granted quite broad protection to the non-literal elements of software and was justified under an incentive theory. Often companies expend a larger part of their dollar investments in software development on design rather than coding.\textsuperscript{170} Interfaces are integral to the design because they dictate how modules com-

\textsuperscript{167} Commentators have debated for years whether copyright, patent, or some hybrid system is the appropriate scheme for protection of computer software. See, e.g., Pamela Samuelson et al., \textit{A Manifesto Concerning the Legal Protection of Computer Programs}, 94 Colum. L. Rev. 2308, 2312 & n.6 (1994) (citing the extensive literature suggesting a sui generis scheme for computer programs and contributing to it). Rather than revisiting that debate, this Article takes the current legal system as given. The chances of implementing a new scheme for software protection are slim. The United States has carried its software protection scheme under patent and copyright law (such as it is) forward into the international arena, limiting its ability to adopt an entirely new system of protection. This Article therefore focuses on improving the current patent construct to address not only problems that may be unique to software, but also market failure in patent law more generally.

\textsuperscript{168} 982 F.2d 693 (2d Cir. 1992).

\textsuperscript{169} See Whelan Assoc., Inc. v. Jaslow Dental Lab., Inc., 797 F.2d 1222, 1226 (3d Cir. 1986) (holding, in a case involving non-literal infringement, that the unprotected idea of a computer program is the overall purpose of the program and stating that "[w]here there are various means of achieving the desired purpose, then the particular means chosen is not necessary to the purpose; hence, there is [protected] expression, not idea"). As the \textit{Computer Associates} court noted, commentators criticized \textit{Whelan} as simplistic, claiming that its failure to recognize that many ideas may exist in one program resulted in overbroad copyright protection for software. See \textit{Computer Associates}, 982 F.2d at 705-06.

\textsuperscript{170} See Whelan, 797 F.2d at 1231 ("By far the larger portion of the expense and difficulty in creating computer programs is attributable to the development of structure and logic of the program, and to debugging, documentation and maintenance, rather
municate within the program and how it connects to other software. A straightforward application of conventional copyright theory would support granting relatively strong rights to interfaces to provide an incentive to invest in their development.

However, by the time of *Computer Associates* in 1992, courts were beginning to understand the economics of software and to realize that broad copyright protection might not be required to provide appropriate incentives. In that case, the Second Circuit, in addressing a claim of non-literal, structural infringement, 171 was the first to label program elements dictated by external factors such as “compatibility requirements of other programs with which a program is designed to operate in conjunction” unprotected by copyright. 172 The application programs at issue necessarily contained some similarities because their design was constrained by the APIs of the third party’s (IBM’s) operating system on which both the plaintiff’s and defendant’s applications were designed to run. If programming code dictated by an operating system’s APIs is uncopyrightable, it is only a small step forward to also hold the APIs themselves uncopyrightable.

The “video game” cases took that step and also used the doctrine of fair use to excuse certain infringement and encourage compatibility. In both *Atari Games Corp. v. Nintendo of America Inc.* 173 and *Sega Enters. Ltd. v. Accolade, Inc.*, 174 independent video game developers (Atari and Accolade) reverse engineered consoles and game cartridges of leading console manufacturers (Nintendo and Sega). The developers conducted the reverse engineering to discover the interface specifications between the consoles and cartridges so they could then produce games compatible with the respective consoles. 175 In both cases, the video game developer

---

171. The claim was not that the infringer copied the source code of the copyrighted product but rather its structure, which “includes its non-literal components such as general flow charts as well as the more specific organization of inter-modular relationships, parameter lists, and macros.” *Computer Associates*, 882 F.2d at 702. The court adopted a three part process for separating a program’s idea from its expression and determining substantial similarity. See id. at 706–12 (explaining the abstraction/filtration-comparison test). After identifying the level of abstractions within a program, see id. at 706–07 (explaining that a program contains not one, but many ideas), the court must filter out the elements that are not protected by copyright. The third step is to compare the protected material to the allegedly infringing work to decide the infringement question: “[T]he court’s substantial similarity inquiry focuses on whether the defendant copied any aspect of this protected expression, as well as an assessment of the copied portion’s relative importance with respect to the plaintiff’s overall program.” Id. at 710.

172. Id.


174. 977 F.2d 1510 (9th Cir. 1992).

175. See *Atari*, 975 F.2d at 836, 842 (describing the steps Atari took in trying to understand Nintendo’s security system, including unlawfully acquiring a copy of source code from the Copyright Office); *Sega*, 977 F.2d at 1514–15 (describing Accolade’s “two-
incorporated code from the manufacturer's system into the game cartridges in order to unlock the console.176

The reverse engineering necessarily involved making a copy of the manufacturers' code as an intermediate step in producing the eventual end-products—compatible games. It was this intermediate copying that was at issue in both cases and that both courts excused as fair.177 The Sega court was more explicitly influenced by the existence of network externalities as it emphasized the public benefit that would result from allowing even unlicensed developers to write compatible games.178 Further, both courts moved beyond Computer Associates by holding that the interfaces themselves were not protected by copyright.179

One way to interpret this latter holding is that it reflected the courts' fears that the console manufacturers were using their copyrights as "disproportionate leverage" into a secondary market.180 Disproportionate leverage cases speak to the impropriety of controlling a standard—and the market(s) it defines—by virtue of one strategically important property right.181 In excusing the copying of short sequences of code to facilitate unlocking the console, the Sega court simply found it untenable that console manufacturers could use copyright to block competition in the after-market for game cartridges.182

A similar rationale could support holding the interfaces uncopyrightable. Unlike patent law, copyright law makes no threshold inquiry to evaluate a work for its contribution to advancing progress. If interfaces that have low intrinsic value are copyrighted, the rightholder may assert market power disproportionate to its technological step forward. This is

---

176. See Atari, 975 F.2d at 836, 845 (noting that after understanding Nintendo's locking program, Atari wrote its own program in another language to perform the same function and finding that Atari's program likely infringed Nintendo's because it contained not simply code to unlock the system but also instructions unrelated to that purpose); Sega, 977 F.2d at 1515–16 (explaining that Accolade incorporated four bytes of data from Sega's system to unlock the console).

177. "The Copyright Act permits an individual in rightful possession of a copy of a work to undertake necessary efforts to understand the work's ideas, processes, and methods of operation. This permission appears in the fair use exception to copyright exclusivity." Atari, 975 F.2d at 842. See Sega, 977 F.2d at 1527.

178. See Sega, 977 F.2d at 1523.

179. Sega so held explicitly. See id. at 1522 ("Accolade copied Sega's software solely in order to discover the functional requirements for compatibility with the Genesis console—aspects of Sega's programs that are not protected by copyright."). Atari's holding to this effect may be inferred from its statement on fair use. See supra note 177.

180. Merges, Charles River Bridge, supra note 77, at 4 (formulating the disproportionate leverage theory).

181. Id. at 8.

182. See id. at 5. In fact, that court went so far as to note that: "When specific instructions, even though previously copyrighted, are the only and essential means of accomplishing a given task, their later use by another will not amount to infringement." Sega, 977 F.2d at 1524 (quoting the CONTU Report at 20).
particularly likely in a network market where competitors (like Accolade) are unable to internalize the network benefit to consumers of increasing the desirability of the rightholder’s product (like Sega’s console). Moreover, the copyright owner’s exercise of its disproportionate leverage may reflect an anti-copyright motive as it seeks to control complementary markets.

The pro-compatibility regime that the preceding cases illustrate has also been enshrined in the Copyright Act. The Digital Millennium Copyright Act (DMCA), enacted in 1998, makes it illegal to circumvent technological devices that block access to copyrighted works or to traffic in devices that would enable such circumvention.\(^{183}\) However, it contains an exception permitting development and use of circumvention tools when reverse engineering is necessary to achieve interoperability with an independently created computer program.\(^{184}\) In other words, the DMCA adopts the rule of Sega.\(^{185}\) At the same time, the common law doctrine of copyright misuse supplements the public laws’ efforts to enable compatibility. Courts can use that doctrine to defeat private efforts to evade the public law through restrictive contractual agreements.\(^{186}\)


\(^{184}\) See 17 U.S.C. § 1201(a)(1), (2). The Act also allows the reverse engineer to circumvent technological devices protecting other rights of the copyright owner, to provide the means of reverse engineering to others to enable interoperability, and defines interoperability. See id. at § 1201(f).

\(^{185}\) Sega had built a lock into its console which allowed access only to cartridges containing the key. See Sega, 977 F.2d at 1575 (describing the TMSS initialization code which had to be contained in a game cartridge in order for it to unlock the console); see also S. Rep. No. 105-190, at 13 (1998) (indicating intent to permit reverse engineering to achieve interoperability as permitted under current law and citing Sega). Despite this exception, it is difficult to know the extent to which the DMCA will encourage interoperability in practice. The interoperability exception may be limited to achieving vertical compatibility. The statute defines interoperability as “the ability of computer programs to exchange information, and of such programs mutually to use the information which has been exchanged.” 17 U.S.C. § 1201(f)(4). Arguably this is broad enough to encompass both vertical and horizontal compatibility. The testimony is conflicting. The Senate Report states that the intent of the reverse engineering exception is to “achieve[e] interoperability to the extent permitted by law prior to the enactment of this chapter. The objective is to ensure that the effect of current case law interpreting the Copyright Act is not changed by enactment of this legislation for certain acts of identification and analysis done in respect of computer programs.” S. Rep. No. 105-190, at 13. At the time of enactment, the “current caselaw” (apart from Lotus Development Corp. v. Baintan Int’l Inc., 49 F.3d 807 (1st Cir. 1995), aff’d by an equally divided Court, 516 U.S. 233 (1996), which did not involve reverse engineering) had primarily fostered vertical compatibility. If a court construes interoperability to exclude horizontal compatibility, the reverse engineer may still be sheltered by the DMCA’s provision that indicates that it has no intent to alter fair use rights. Finally, note that the DMCA’s otherwise broad ban on trafficking in anti-circumvention devices may make it more difficult for the reverse engineer it purports to shelter to obtain a device to break a technological lock designed to protect an interface.

\(^{186}\) In the case most closely on point, Alcatel USA Inc. (DSC) v. DGI Tec, Inc., 166 F.3d 772 (5th Cir. 1999), the Fifth Circuit upheld a finding of copyright misuse where the licensor attempted to use its copyright as leverage in the market for compatible products. DSC markets telephone-switching systems that include copyrighted operating system
But withdrawing copyright protection for interfaces leaves open the question of how best to provide an incentive for the production of truly innovative interfaces. Without much discussion, the Atari and Sega courts suggested that the console manufacturers should have met the standards of patent law to obtain the broad rights they were seeking. Because both Acolade and Nintendo sought vertical compatibility that would actually enhance the demand for the copyrighted works of Sega and Atari, the incentive question seemed somewhat irrelevant. In contrast, in *Sony Computer Entertainment, Inc. v. Connectix Corp.*, the court addressed the question of horizontal compatibility. There, defendant Connectix reverse engineered Sony's console video game system (the PlayStation). The PlayStation allows a user to play video games on his or her TV screen. Connectix developed a program that would run PlayStation games on a PC. Applying Sega, the Ninth Circuit held the reverse engineering to be fair use despite its resulting in an eventual end-product (the Virtual Game Station) that is a direct competitor of, and, if successful, could be a market replacement for Sony's PlayStation (in which Sony had invested over $500 million).

software, switches and microprocessor cards. See id. at 777–78. During its operation, the microprocessor card downloads a copy of the operating system into RAM. See id. at 778. DSC licensed the operating system under terms restricting the licensee from copying it and using it in conjunction with other manufacturers' equipment. See id. at 777. The court held the licensing arrangement to be a misuse of copyright since:

any competing microprocessor card developed for use on DSC phone switches must be compatible with DSC's copyrighted operating system software. In order to ensure that its card is compatible, a competitor such as DGI must test the card on a DSC phone switch. Such a test necessarily involves making a copy of DSC's copyrighted operating system . . . . If DSC is allowed to prevent such copying, then it can prevent anyone from developing a competing microprocessor card, even though it has not patented the card.

Id. at 793–94 (citing DSC Communications Corp. v. DGI Techs., Inc., 81 F.3d 597 (5th Cir. 1996).

Similarly, in *Practice Management Info. Corp. v. American Med. Ass'n (AMA)*, 121 F.3d 516 (9th Cir. 1997), the Ninth Circuit held that the AMA's license requiring exclusive use of its copyrighted coding system was copyright misuse because it gave the AMA an unfair advantage in the marketplace by restricting the licensee's choice in using a competitor's product. See *Practice Management*, 121 F.3d at 521 (“What offends the copyright misuse doctrine is not [the licensee's] decision to use the AMA's coding system exclusively, but the limitation imposed by the AMA licensing agreement on [the licensee's] rights to decide whether or not to use other forms as well . . . . The [license] terms . . . gave the AMA a substantial and unfair advantage over its competitors.”). See also supra note 61 (describing the *Lasercard* case).

187. See Atari, 975 F.2d at 842; Sega, 977 F.2d 1510, at 1525.
188. 203 F.3d 596 (9th Cir. 2000).
189. See id. at 598.
190. See id. at 602–08

Because the Virtual Game Station is transformative, and does not merely supplant the PlayStation console, [it] is a legitimate competitor in the market for platforms on which Sony and Sony-licensed games can be played . . . . [S]ome economic loss by Sony as a result of this competition does not compel a finding of no fair use. Sony understandably seeks control over the market for devices that
Thus horizontal compatibility starkly raises the incentive question. Why should Sony invest in development if second comers can free-ride on its investment, undercut its price, and usurp its market? The court answered this question by directing Sony to patent law, stating, "If Sony wishes to obtain a lawful monopoly on the functional concepts in its software, it must satisfy the more stringent standards of patent laws." At least in the Ninth Circuit then, the movement of intellectual property protection for interfaces from the copyright law to patent law, whether to facilitate vertical or horizontal compatibility, is complete.

The concurrence in another horizontal compatibility case, *Lotus Development Corp. v. Borland Int'l Inc.*, in the First Circuit, shows how fair use can profitably tailor rights over time, answering incentive concerns. The majority used Section 102(b) of the Copyright Act to enable horizontal compatibility in the applications market, holding Lotus's 1-2-3 spreadsheet menu command hierarchy an uncopyrightable method of operation. Using the rhetoric of network effects, the court emphasized that users invest time and money in learning how to operate a system and store files in its format. If the customer cannot transfer its investment to another system, he or she will not incur the costs to switch even to a superior system. The court found this possibility "absurd." Professor Merges characterizes this view as a concern over a particular species of disproportionate leverage that he terms "user holdup"—the idea that some standards succeed not so much because of the work's merit but because of the collective investment of purchasers.

However, in his concurrence, Judge Boudin noted how difficult it is to define the scope of copyright protection for computer software to provide appropriate incentives to innovate. As he indicated, the essential play games Sony produces or licenses. The copyright law, however, does not confer such a monopoly.

Id. at 607.

191. The current emphasis on sheltering primarily "transformative uses" as fair, see supra note 58 (describing the "transformative use" strand of fair use jurisprudence), makes it more likely (despite Sony) that a court will excuse infringement where the end result is a vertically rather than horizontally compatible product. However, the presence or lack of a transformative use is not determinative—as Sony suggests—leaving a court with leeway to excuse even infringement resulting in horizontal compatibility when appropriate.

192. Id. at 605.

193. 49 F.3d 807 (1st Cir. 1995), aff'd by an equally divided Court, 516 U.S. 233 (1996). Other courts have held similarly. See Menell, supra note 139, at 702–03 (identifying and reviewing decisions that relate to command structures and command codes).

194. See *Lotus*, 49 F.3d at 815. However, the court expressly did not address the question of whether Borland had infringed Lotus's audiovisual copyrights. See id. at 812 (noting that Lotus did not appeal the question of whether Borland copied its screen displays).

195. See id. at 817–18.

196. Id.

197. Merges, Charles River Bridge, supra note 77, at 8–9 (indicating that in some such cases, the creator's rights may appropriately be subordinated to those of users).
problem is the same for software as for other copyrighted works—how to encourage creativity without unnecessarily constraining access.\textsuperscript{198} The difference is that in the case of software, the cost of a mistake in granting too much protection may be much higher because of network effects.\textsuperscript{199}

Judge Boudin also emphasized the temporal nature of the inquiry, noting that Lotus had become the de facto standard spreadsheet, rendering it difficult for another program to compete unless it could offer compatibility.\textsuperscript{200} In other words, the market was characterized by stubborn customer expectations and entrants could not compete solely on traditional metrics like quality and price. At the same time, “Lotus had already reaped a substantial reward for being first.”\textsuperscript{201} The inference is that Borland would not have innovated at all if it could not copy Lotus and Lotus would still have done so even had it known that it would be forced to give Borland access after the 1-2-3 spreadsheet program became dominant.

Judge Boudin noted that fair use could be an alternative justification for the court’s holding.\textsuperscript{202} As in the video game cases, the externality and anti-copyright motives justifying a fair use finding might be implicated. Moreover, a fair use approach could have been more effective in safeguarding innovation incentives. By holding the menu command hierarchy uncopyrightable, the court gave blanket protection to the slavish copier as well as to someone like Borland who added substantial value to

\begin{flushleft}
198. See \textit{Lotus}, 49 F.3d at 819 (Boulin, J., concurring).
199. See id.

The problem presented by computer programs is fundamentally different in one respect. The computer program is a means for causing something to happen; it has a mechanical utility . . . . Of course, the argument for protection is diminished, perhaps even enhanced, by utility: if we want more of an intellectual product, a temporary monopoly for the creator provides incentives for others to create other, different items in this class. But the ‘cost’ side of the equation may be different where one places a very high value on public access to a useful innovation that may be the most efficient means of performing a given task. Thus, the argument for extending protection may be the same; but the stakes on the other side are much higher.

Id.

200. See id. at 820–21.

Apparently, for a period Lotus 1-2-3 has had such sway in the market that it has represented the de facto standard for electronic spreadsheet commands. So long as Lotus is the superior spreadsheet—either in quality or in price—there may be nothing wrong with this advantage. But if a better spreadsheet comes along, it is hard to see why customers who have learned the Lotus menu and devised macros for it should remain captives of Lotus because of an investment in learning made by the users and not by Lotus.

Id.

201. Id. at 821.

202. See id.; see also David R. Owen, \textit{Note. Interfaces and Interoperability in \textit{Lotus v. Borland: A Market-Oriented Approach to the Fair Use Doctrine}}, 64 Fordham L. Rev. 2381, 2400–04 (1996) (arguing that the First Circuit was incorrect in holding the menu command hierarchy uncopyrightable and that fair use can be used to promote interoperability).
Lotus's work. While "compatibility" includes both horizontal and vertical compatibility, the incentive effects of opening the standard to competition vary in each instance. The fair use doctrine offers a way for a court to tailor incentives more easily than a bright-line holding on whether a work is or is not protected by copyright.

Juxtaposing the video game cases with Judge Boudin's concurrence in Lotus raises an important question. The video game cases made interfaces the subject matter of patent law rather than copyright. However, Judge Boudin points out that had they remained the subject matter of copyright, that law's fair use defense might profitably have fine-tuned the scope of the rights over time. Might an analogous doctrine therefore also perform a useful function in tailoring the scope of API patents?

This is no mere theoretical question. Microsoft alone has rights to at least twenty-two patents relating to APIs. Indeed, Sony apparently also owns such patents. Shortly after the Ninth Circuit rejected its copyright claim, Sony sued Connectix for patent infringement. This case may force a court to consider the consequences of relegation of API protection to patent law. This or another court in the near future will have to decide whether the policy arguments in favor of a compatibility regime in copyright resonate in patent law. The following part of this Article argues that they do and that the new patent fair use defense could be

---

203. A 1997 search by Professors Lemley and McGowan "revealed 22 patents assigned to Microsoft which dealt with APIs." See Lemley & McGowan, supra note 138, at 529 n.212. This is not to argue that Microsoft's allegedly dominant position in the operating systems market is attributable to its ownership of patents. Software has only recently been viewed as patentable. Microsoft's market power is more likely attributable in part to its copyrights, industry uncertainty (until recently) as to whether interfaces are copyrighted, and the sheer technical difficulty of replicating the Microsoft system. At this time, the patents would be useful in maintaining Microsoft's market share. The question is whether, given its antitrust difficulties, Microsoft would choose to assert them aggressively.


205. One question is why such cases have not arisen before. Traditional hardware markets involve interfaces that allow devices to connect and patents have played a key role in such industries. However, there is a dearth of caselaw involving a defendant seeking the right to infringe either to replicate an interface or to develop a product which conforms to it. Professor Merges offers some plausible reasons why this might be the case. See Merges, Charles River Bridge, supra note 77, at 17-18 (noting that companies often own patents on components of a product such that a number of firms must—and do—cross license in order to produce and sell the product). There were some cases brought against IBM in the 1960s and 1970s that dealt with the question of whether IBM could change its interfaces and thereby render third party products incompatible. These cases were grounded in antitrust claims, with no mention of patent law or whether the interfaces were protected by patent rights. In them, the courts rather uniformly held that IBM could change its interfaces even if the effect were to render other vendors' products incompatible. However, the decisions contained some warnings. See, e.g., In re IBM Peripheral EDPL Devices Antitrust Litig., 481 F. Supp. 955, 1002-03 (N.D. Cal. 1979) (aff'd sub nom., Transamerica Computer Co. v. IBM, 695 F.2d 1377 (9th Cir. 1983)) (too-frequent changes of interface for no purpose other than preclusion of competition might constitute predatory behavior); Kevin J. Arquité et al., Antitrust, Intellectual Property, Standards and
deployed to open APIs to competition through a compatibilty regime when appropriate.

3. Patent Law and Compatibility rules. — In a network market, the API, whether or not "obvious" in a patent sense, will owe much of its value to the number of consumers that adopt it rather than to its intrinsic worth. A patented interface may become a standard, leading to undesirable double dominance over product-market and R&D competition. While a firm may require an incentive to invest in the non-obvious, patentable interface, it is unclear to what extent broad patent rights enabling this double dominance are required to provide that incentive. In fact, social welfare may be maximized by opening the standard to competition under circumstances in which it is reasonable to conclude that the patentee's incentives will not be unduly harmed. Patent fair use could usher in such beneficial competition.

For example, if a software firm owns a patent on its operating system API, no competitor will be able to offer a horizontally compatible product without infringing it. Additionally, no unlicensed application developer is likely to be able to write a vertically compatible application without reverse engineering the operating system to gain access to the APIs and other information,\textsuperscript{206} a practice that is likely to constitute patent infringement.\textsuperscript{207} But it may be socially desirable to open the API to competition at some point.

Interoperability, 524 PLI/Pat. 157, '85-86 (1998) (reviewing antitrust cases involving changes to product design).

\textsuperscript{206} The need to reverse engineer to obtain the interface specification is a function of how much detail is disclosed by the patentee in the enabling disclosure required by statute as a condition of patentability. See 35 U.S.C. § 112 (1994 & Supp. III 1997). The more detail, the less the need to reverse engineer. However, in the area of software patents, the PTO has not demanded a particularly painstaking level of disclosure. See, e.g., Michael J. Walsh, Comment, The Disclosure Requirements of 35 U.S.C. § 112 and Software-Related Patent Applications: Debugging the System, 18 Conn. L. Rev. 855, 856 (1986) (contending that software disclosure requirements are insufficient and suggesting changes); see also Lemley & McGowan, supra note 138, at 524 & n.195 (noting that reverse engineering may infringe a patent and stating "[w]hile in theory the disclosure required by 35 U.S.C. § 112 should reduce the need to reverse engineer a patented product, in practice there are several reasons why a competitor might need to reverse engineer a patented product. First, she may need access to unpatented components of the same device. Second, the patent specification may not describe the invention in sufficient detail. Finally, compatibility may require an exact picture of the interfaces between two products, and thus may require more detailed information than that the patentee is forced to disclose.").

\textsuperscript{207} The legal status of reverse engineering is debatable. See Julie E. Cohen & Mark A. Lemley, Patent Scope and Innovation in the Software Industry 13-14 (March 24, 2000 draft, on file with the Columbia Law Review) ("Reverse engineering a patented computer program likely falls within [the] broad category of prohibited conduct in at least some cases . . . . At the very least, reverse engineering constitutes a 'use' of the patented software . . . . [and] most likely it also constitutes 'making' the patented invention . . . . And certain reverse engineering involves the making not just of copies of the software in RAM, but also longer-term . . . . copying. Those copies probably constitute patent infringement, unless protected by some defense."). However, they also note that the common law patent
In any given case, a court might effectively open an API to competition by holding the patent encompassing it to be invalid or by constraining it narrowly to avoid holding the infringer liable. It seems unlikely that a court would hold APIs globally not to be within patent law’s purview. Courts have broadly interpreted the subject matter of patent law to encompass “anything under the sun that is made by man.” Moreover, investors may require some incentive to invest in production of innovative APIs. Because copyright has delegated the field to patent, a withdrawal of such protection could seriously undermine production incentives. The real question is how to define the patent rights to avoid providing too much of an incentive.

If the patent is valid and the infringing invention reads on to its claims, a court might use the reverse doctrine of equivalents to excuse the infringement. Recall that that doctrine allows a court to excuse the infringement of a radical improver, helping to overcome licensing holdup problems like those that may characterize network markets. Moreover, this radical improvement may help to displace the market power a single firm standard owner has attained.

But, in most instances, the reverse doctrine of equivalents will not offer much comfort to the infringer. Regardless of whether the end-product is horizontally or vertically compatible with the patented one, the reverse engineering that constitutes infringement is not a radical improvement. Further, if the infringer’s eventual end-product were a hor-

doctrines of implied license and exhaustion support a right to reverse engineer. See id. at 21–25. The Supreme Court, in explaining why state trade secret law may co-exist with federal patent law, has suggested that the reverse engineer would be an infringer: “In essence the . . . law [at issue in this case] prohibits the entire public from engaging in a form of reverse engineering of a product in the public domain. This is clearly one of the rights vested in the federal patent holder, but has never been a part of state protection under the law of unfair competition or trade secrets.” Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 164 (1989). The Sony case may resolve this question, although it is unclear from the complaint whether or not Sony is challenging Connectix’s reverse engineering. See Sony Computer Entertainment v. Connectix Corp., No. 00-CV-520, ¶ 9, (N.D. Cal. filed Feb. 14, 2000). Note also that depending on how the claims are written, the application itself that uses the API may also infringe the patent if the rightholder has patented the process of using the API with an application program. The Sony case may involve such patents. See id. at ¶ 9 (listing allegedly infringed patents that include a number of “method claims”). If the patent is on the API itself, the end user is in the position of a direct infringer but is likely sheltered by the first sale or patent exhaustion doctrine. In the absence of direct infringement, the application provider cannot be liable for contributory infringement. See Arco Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 483 (1964) (citing Arco Mfg. Co. v. Convertible Top Replacement Co., 365 U.S. 396, 401–402 (1961)). The case for fair use does not rest on the legality of reverse engineering. While fair use may be useful in addressing such conduct, its primary utility will occur in evaluating conventional infringements and infringing end products.

208. Diamond v. Diehr, 459 U.S. 175, 182 (1981) (citation omitted). However, certain items, including “laws of nature, natural phenomena, and abstract ideas,” are unpatentable. Id. at 185.

209. See Lemley, Improvement, supra note 10, at 1067 (suggesting that the reverse doctrine of equivalents offers a possible solution to second comers in a network market).
zontally compatible operating system, that too would likely not be a radical improvement. To write a compatible operating system, the secondcomer must implement the same APIs to run the applications in which users have invested. Radically improving them is simply not a viable option when the firstcomer has a stubborn base of customers who have installed and are running its system.

However, the reverse doctrine of equivalents might assist the application developer if it could convince a court to adjust the doctrine to resemble copyright fair use as it was applied in the video game cases like Atari, Sega, and Sony. The application itself would be a new work, performing a function substantially different from the operating system. As courts have used copyright fair use to excuse the "intermediate" infringement that occurs in the production of a new, complementary product, so too might they use the reverse doctrine of equivalents to excuse "intermediate" reverse engineering to produce a new, interoperable product.

If the infringer improves the patented invention after the reverse engineering, it may itself obtain a blocking patent. Again though, it may be difficult for the infringer to improve the invention and simultaneously maintain compatibility. Even if it could, the usual assumption that the original patentee will negotiate with the infringer for access to the improvement (and to excuse the reverse engineering) may not hold. A patentee who seeks to control innovation in the primary and secondary markets may prefer to block the secondcomer from market entry rather than to engage in a cross-license. Also, in an industry like software that is characterized by incremental improvements and constrained by backward compatibility concerns, the secondcomer’s improvement itself may not rise to the level of patentability, rendering the doctrine of blocking patents inapplicable. Nor would the experimental use exception excuse the infringement, because of its ultimate commercial purpose.

Finally, patent misuse is also unlikely to assist the infringer. The patent misuse doctrine is a bit amorphous, tending to wax and wane as attitudes toward patent enforceability shift over time. A contractual term

210. See supra notes 173–179 and notes 188–190 and accompanying text (discussing the video game cases).

211. Backward compatibility refers to the practice of safeguarding user investments by ensuring that existing applications running on an operating system today will also run on updated versions of the operating system tomorrow. In the applications market, it refers to the practice of releasing new versions of applications with the ability to process files saved using the older version. For example, Word Version 6.0 can process files saved in Word Version 2.0.

212. Both the horizontal competitor and the application developer plan to market a product either in direct competition with or in the after-markets for the patented product. See Grossman, supra note 15, at 244 (noting that the experimental use cases may excuse defendants “when the use of the patented invention is for the sole purpose of gratifying curiosity or a philosophical taste, or for mere amusement”); supra notes 69–70 and accompanying text.

213. See Merges, Charles River Bridge, supra note 77, at 26–27. However, Professor Merges has identified the same theme of "disproportionate leverage" that helps to explain
that prohibits reverse engineering of a patented product would simply
restate rights the patentee already has. Arguably, it overreaches because
it prevents the licensee from obtaining other, non-patented information.
But even if a court were to allow reverse engineering to obtain such in-
formation, it would be unlikely under conventional law to accord the reverse
engineer the right to obtain and use the patented APIs in a commercial
end-product.

4. Applying Patent Fair Use to APIs. — Fair use could assist a court in
evaluating the propriety of infringing an API patent. This approach
would provide the court with the doctrinal latitude to consider relevant
policy concerns and would not require contortion of existing patent
doctrine to achieve a desirable result.

a. Fair Use. —

i. Nature of the Advance. — In assessing this first fair use factor, a
court would likely cast a more favorable view toward the indirect infringe-
ment conducted in writing a compatible application than the directly in-
fringing horizontally compatible operating system that may result from
reverse engineering. The application is, in copyright language, a “trans-
formative” use, potentially adding substantial value to the infringed oper-
ating system. In contrast, a competing operating system does not advance
the art, although it is possible that the infringer could innovate within the
system, making it run more efficiently. Generally therefore, this first fac-
tor would weigh in favor of the application developer and against the
operating system provider. However, it must be considered in light of the
other considerations.

ii. The Purpose of the Infringing Use. — Under this second factor, the
court should consider the use to be commercial whether the infringe-
ment is conducted to market a directly infringing operating system or as a
step in producing an application. While neither should bar a finding of
fair use, horizontal compatibility, which enables replacing the patented
work in the marketplace, is more likely to affect the patentee’s original
incentives adversely than vertical compatibility, and therefore should be
more carefully scrutinized.

Additionally, there is a distinction between these examples and the
copyright cases. The copyright courts in Atari, Sega, and Sony excused
copying involved in reverse engineering because it was required to obtain
access to elements the copyright law did not protect. But here, the re-
verse engineering is required to obtain the patented invention itself. It is
therefore not directly analogous to that involved in the copyright cases
and is more threatening to the core property right to exclude.

the copyright video game cases as accounting for at least some subset of patent misuse
jurisprudence. See id. at 26. He suggests that courts could use patent misuse in
conjunction with the reverse doctrine of equivalents to adopt pro-compatibility rules like
those of copyright, although he acknowledges that these doctrines may be useful primarily
"at the margins." Id. at 28–31 (giving an example of how misuse could address situations
in which the patentee has bundled patented software with another application).
Again, that threat is particularly obvious when the end product is a horizontally compatible operating system whose commercial success is a direct product of infringing the patent. In contrast, the success of the application program that does not itself infringe the patent is only indirectly attributable to the infringing activity. Thus, the second factor may be neutral with respect to the application developer but would weigh against the purveyor of the competitive operating system.

iii. Market Failure. — The third and fourth factors are the most important in addressing a fair use claim. In weighing the third factor—the nature of the market failure that frustrates licensing in the particular case—courts must seek to understand the patentee’s licensing practice. For example, a refusal to license application developers is more suspect than the refusal to license the maker of a competing operating system. Failure to license application developers is both particularly troubling and likely to occur when the patentee also competes in the application market and seeks to give its own applications a competitive advantage.

In the operating system context, because the marginal cost of producing an additional copy of patented software is near zero, the patentee can meet all of the market demand without needing to license the invention to others. Of course, this effectively locks users into a single supplier. When the court believes the social benefit of enabling competition outweighs the harm to the patentee’s incentives, it should be willing to open the interface even to horizontal competitors.

This third factor inquiry is delicate because the exclusive right of the patent includes the patentee’s right not to exploit it, either independently or by licensing it to others. In the API example, the patentee has chosen to exploit the invention, but not to license it out. If the law is interpreted to require that the owner of the dominant API must license it, is the legal rule effectively undercutting the right not to license and imposing a penalty for success?

American law has never been interpreted to allow a rightholder to extract all of the profits it might be able to glean from exploitation of its invention. In particular, the price of rampant success has always been an increased social responsibility. This reality is reflected in statutory constructs like “genericide” in trademark law, the antitrust doctrine of essential facilities and the rule that restrictive conduct, permissible in

---


215. Under the Lanham Act, a generic mark may not be registered and even an incontestable mark may be cancelled if it becomes generic. See, e.g., 15 U.S.C. § 1064 (1994) (providing for cancellation of a mark if it becomes the “generic name for the goods or services . . . for which it is registered”). This doctrine of genericity helps to “adjust the [trademark] incentive to call forth the optimum (or at least desirable) level of this activity, rather than providing a permanent, blanket incentive.” Merges, Charles River Bridge, supra note 77, at 37.

216. Under the essential facilities doctrine, a monopolist who controls the gateway to a second market may be required to grant competitors access to that facility. See MCI Communications Corp. v. American Tel. & Tel. Co. 708 F.2d 1081, 1132-33 (7th Cir. 1983).
a competitive market, may be illegal when a monopolist engages in it. Under the genericide doctrine, the trademark owner loses its rights completely when it is so successful that the public appropriates its mark as the name of the product. Under the essential facilities doctrine, the property owner is required to provide others a privilege of access and use to safeguard competition in secondary markets. The rule constraining a monopolist's behavior effectively places more responsibility on the monopolist to deal fairly with the competition than it would have if it lacked market power.

In patent law, the right not to license is enshrined in the Act itself, which provides that a finding of patent misuse may not be based on "refus[al] to license or use any rights to the patent." Nevertheless, even patent courts have been troubled by such refusals when the public welfare is at issue. Fair use could supplement misuse by giving a court the authority to order a license when necessary to protect the public or advance its welfare. A court should be more willing to do so when the patentee has exploited the patent commercially than when it has chosen (setting forth elements of an essential facilities claim). However, that doctrine has never been used to require an intellectual property rightholder to grant a license. See Intergraph Corp. v. Intel Corp., 195 F.3d 1346, 1356-58 (Fed. Cir. 1999) (overruling a lower court ruling that employed the doctrine to require Intel to provide information protected by intellectual property rights to Intergraph, but basing its ruling not on intellectual property grounds but rather on the lack of a competitive relationship between the parties).

217. See, e.g., United States v. Aluminum Co. of Am., 148 F.2d 416, 432 (2d Cir. 1945) (noting that the plaintiff had satisfied the court that it was unnecessary to demonstrate "practices unlawful of themselves" to sustain a monopolization claim).

218. 15 U.S.C. § 271(e)(4) (1994). Subsection (e)(4) was added at the same time as (e)(5). Subsection (e)(5) provides that a finding of misuse cannot be based on tying rights to the purchase of an unpatented product unless the patentee has market power. See id. at § 271(d)(5). Despite the simultaneous addition of these subsections, the plain language of the statute indicates that the possession of market power does not also allow a court to find misuse because of a unilateral refusal to license. But see Schlumberger v. Public Key Partners and RSA Data Security, Inc., 1997 U.S. Dist. LEXIS 15287, *30 (reading § 271(d) as authorizing a finding of misuse for a refusal to license where the defendant has market power). One question is whether § 271(d)(4) addresses only cases of patent suppression or also cases where the patentee uses the invention but does not license it to others. The statutory language is broad enough to cover both. See Jay Drialer, Jr., Licensing of Intellectual Property § 3.92 (1999) (claiming that § 271(d)(4) 's wording "makes clear beyond doubt that neither refusal to license nor suppression of patents constitutes patent misuse").

219. See, e.g., Vitamin Technologics, Inc. v. Wisconsin Alumni Research Found., 146 F.2d 941, 944, 956 (9th Cir. 1945) (expressing concern over the patentee's refusal to license a patent that could help to make margarine useful in preventing violets in order to safeguard the dairy industry. "It is now well established that a patentee may not put his property in the patent to a use contrary to the public interest... Upon consideration of the proffered evidence, all three judges now conclude the refusal [to license] was warranted and against the public interest."); see also supra note 130 and accompanying text (discussing the Activated Sludge case).

220. See also supra note 121 (discussing the Biley database protection bill that considers licensing conduct in assessing misuse).
not to avail itself of the benefits and concomitant responsibilities of the marketplace.

iv. Impact on Incentives and Social Welfare. — The facts relevant to the third factor also play a role in the fourth—the market harm to the patentee weighed against the social benefit. There seems to be relatively little harm where the infringer markets a vertically compatible application because that product enhances the demand for the patented one and benefits society by increasing the value of the network. If the reverse engineering is truly necessary to produce that vertically compatible product, why should the patentee be able to enjoin it? The patentee has no incentive to discourage products that augment the demand for its own unless it wishes to leverage its operating system's power into the application market, controlling who competes there as well as in the primary market where it has patent protection. Such behavior evinces an anti-patent motive that fair use might be deployed to defeat.221

In contrast, the market harm to the patentee of a directly infringing competing operating system is obvious and may be substantial. Again, the distinction from copyright is striking. In Sony, the court allowed horizontal compatibility because the new market entrant was offering a product that did not infringe Sony's copyright. It merely implemented the uncopyrightable APIs. The court there explicitly directed Sony to the patent law as the source of rights that could be exercised to prevent such conduct from occurring. From this one can infer that, at least in the opinion of the Sony court, patent law would not and should not allow the marketing of a product that directly infringes patented APIs.

However, economic analysis teaches that in network markets, a horizontal compatibility regime may be desirable at certain points in time. Thus, rather than rejecting a fair use claim on its face when horizontal compatibility is at issue, a court should assess whether, in fact, network effects are present as well as their nature and strength. It should be more receptive to a fair use defense when the network effect is strong, producing entrenched consumer expectations and a durable standard from which others are excluded,222 and when it, like the Lotus court, is confident that the patentee has already reaped a "substantial reward" from the

---

221. The patentee may argue that unlicensed applications may erode its goodwill as users may attribute performance problems to the underlying operating system rather than to the application. But see Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 609 (9th Cir. 2000) (refusing to uphold a trademark tarnishment claim when the defendant's games did not play as well on the Sony system as Sony's own games, noting that it was not persuaded "that the difference in quality between the two platforms is itself sufficient to find tarnishment," but also indicating it might have held otherwise if more evidence were presented).

222. See Lemley & McGowan, supra note 138, at 593 ("Because strong network effects tend to produce more uniform standards with more durable effects, the law should be more concerned with attempts to control or exclude others from participating in a standard if the standard is part of an actual or strong virtual network than if the battle is merely over convenience effects.").
innovation. Then, the social benefit of increasing innovation and product-market competition may outweigh the cost to the patentee of allowing the infringement.

v. **Subject Matter.** — In considering the fifth factor, courts should assess whether the API represents a major step forward or whether its value derives more from the first mover advantages and accident of consumer acceptance that characterize a network market. In the latter case, the likelihood that the patentee is exerting “disproportionate leverage” in excluding both horizontal and vertical competitors and locking consumers into one supplier is much higher. In such a case, this factor should weigh in favor of both the application and operating systems developers.

The outcome of the fair use defense is difficult to predict because much hinges on the third and fourth factors. The application developer is likely to have a good fair use claim if market defects are preventing licensing. Since factors one and two weigh against the operating system developer, it has a more difficult hurdle to overcome under the remaining steps of the inquiry. But a court should excuse even this infringement as fair if it is convinced that market defects exist such that licensing will not occur and any harm to the patentee’s incentives is outweighed by the social benefit.

b. **Free or For a Fee?** — While it may be appropriate policy to excuse an infringement as fair under the five-factor test, this does not answer the question of whether the infringement should proceed without any compensation to the patentee.223 When the patentee is refusing to license because it has an anti-patent motive, it may be appropriate not to require payment as a penalty for this behavior. But when the problem is more one of externalities, awarding a royalty may both help to enable competition and safeguard patentees’ incentives. Since the API patents are likely owned by one vendor, the transaction costs of setting up a payment system should be relatively low enough to make compensation feasible.

However, that compensation should be set at something less than fair market value to provide the appropriate incentives.224 Reverse engineering is a time-consuming, difficult, and sometimes inaccurate process.225 Moreover, the patentee may enhance its interfaces at any time, making it likely that the competitor who depends on reverse engineering

223. Cf. Samuelson et al., supra note 167, at 2414-15 (discussing an option for a sui generis software protection regime that would include a period of anti-cloning protection followed by a compulsory license).

224. Cf. F.M. Scherer, The Pharmaceutical Industry and World Intellectual Property Standards, 6-7 (2000) (draft, on file with the Columbia Law Review) (noting that where the government infringes patents in the national interest, the requirement that it pay "reasonable compensation" is interpreted to mean "much lower payments than the foregone monopoly profits standard imposed in private patent infringement damages cases").

225. See Microsoft Findings of Fact, supra note 136, at ¶ 52, 77 (describing "cloning the thousands of APIs already exposed by Windows [as] an enormously expensive undertaking"); Lemley & McGowan, supra note 138, at 529 (noting the technical
will always lag behind the patentee in the marketplace as it struggles to keep up with these changes and the applications that take advantage of them. 226

A fair use right to reverse engineer and, under certain conditions, to infringe the patent by marketing a horizontally compatible product does not solve this problem. Practically, a market entrant is most likely to rely on reverse engineering alone when the operating system (like those in the video game cases or smaller devices like cellular phones) is relatively small, containing a manageable number of APIs. Frankly, nothing short of a far more radical rule that would mandate ongoing pre-market disclosures of interface changes will ensure that others besides the standard owner will be able to compete effectively. But such a rule would largely eradicate the incentive effect of patent protection.

However, even in the case of more complex systems, patent fair use would give the infringer a threat to use in negotiations, increasing the chances that the parties will reach a socially beneficial licensing agreement. 227 By placing the compensation rate, if any, at something less than fair market value, the patentee is less likely to view competition through reverse engineering as an empty threat. It therefore may be more amenable to entering into a royalty-generating license. The infringer is also likely to prefer a license to enable it to compete in a timely manner. Additionally, the “infringer” as licensee might then also obtain access to other, unpatented information that would help to make its development effort less costly. Yet, if the patentee refuses to license, the fair use doctrine could act as a safety valve, excusing the infringement when it would be socially beneficial.

---

226 See Microsoft Findings of Fact, supra note 136, at ¶ 52, 77 (“More daunting [to a competitor] is the fact that Microsoft continually adds APIs to Windows through updates and new versions. By the time a rival finished cloning the APIs currently in existence, Windows would have exposed a multitude of new ones. Since the rival would never catch up, it could never be able to assure consumers that its operating system would run all of the applications written for Windows.”); Lemley & McGowan, supra note 138, at 529 (noting that Microsoft often changes its operating system and competitors would have to keep up with such changes). This problem is mitigated by concerns of backwards compatibility. Normally, any upgrades to the operating system will continue to run applications written to the interfaces of an earlier system, rendering wholesale changes to existing interfaces on a regular basis rather unlikely. See supra note 211. Additionally, the dominant provider would have to convince consumers to upgrade to the enhanced system.

227 See generally Ian Ayres & Eric Talley, Solomonic Bargaining: Dividing a Legal Entitlement To Facilitate Coasean Trade, 104 Yale L.J. 1027 (1995) (discussing the benefits of divided entitlements in facilitating trade by inducing information exchange). Cf. James B. Kobak, Jr., Intellectual Property Law, Competition Law and Hidden Choices Between Original and Sequential Innovation, 3 Va. J.L. & Tech. 6, ¶ 28 (“Broad and widespread intellectual property protection . . . leads to uncertainty and risk for infringers; since an innovator is likely to be both a holder of its own rights and a potential infringer of someone else’s, uncertainty is likely to create incentives for widespread licensing.”).
B. Applications of Fair Use in Other Industries

As a fact-intensive test, fair use is likely to be quite expensive. Its implementation in patent law is justified only if its benefits exceed its costs. This is somewhat unlikely if its utility is confined to the limited and relatively unusual context of patented APIs. However, fair use is likely also to play a valuable role in non-network industries that suffer from their own market defects.

For example, as the subject matter of patent law has expanded to encompass previously unpatentable items and new technologies, rights have tended to splinter, requiring a patentee to seek licenses from a number of other rightsholders in order to market a product. The high transaction costs of seeking these multiple permissions may discourage invention in the first place. While one way to decrease these costs is for rightsholders voluntarily to pool their patents, this is not a viable option if the conditions that facilitate this type of collective action do not exist. A fair use doctrine would help to break logjams in licensing negotiations when rights are widely distributed, mitigating not just transaction-cost problems, but also blockages caused by strategic bargaining.

Professors Michael Heller and Rebecca Eisenberg have identified biotechnology as an area in which the proliferation of intellectual property rights among a number of patentees may create the tragedy of the "anticommons." In an anticommons, all owners have a right to exclude and none has a privilege of unlicensed use, leading to under-utilization of resources. As Professor Heller notes, unlike copyright law, which addresses the anticommons problem through fair use, patent law historically "has [had] few formal mechanisms to deter excessive fragmentation." Instead, it has relied on rightsholders to coalesce in informal institutions to overcome transaction costs.

228. See supra note 282 (citing the State Street Bank case for eliminating the "business methods" exception to patentability).

229. See, e.g., Diamond v. Chakrabarty, 447 U.S. 303 (1980) (upholding a patent on a genetically engineered bacteria). The Chakrabarty decision helped biotech start-up firms to attract venture capital by providing some measure of certainty that biotechnological inventions could be patented. See Kitch & Perlman, supra note 19, at 840.


231. See id.

232. Michael A. Heller, The Boundaries of Private Property, 108 Yale L.J. 1163, 1175 & n.61 (1999) ("Compared with patent law, copyright law's tragedy of the anticommons is less costly. The 'fair use' doctrine means that copyright holders do not have the right to exclude nonowners from low-intensity uses of protected works.").

233. See Heller & Eisenberg, supra note 230, at 700 (noting that repeat players sometimes reduce transaction costs by voluntarily forming licensing institutions, giving the music, automobile, aircraft manufacturing, and synthetic rubber industries as examples, but noting that some patent pools have come about at the government's behest); see also supra note 82 (setting forth citations to the same effect).
However, for several reasons, rightsholders have yet to come together to form a licensing institution in the biomedical field. In the biotechnology and pharmaceutical areas, patents have traditionally played an important role in securing market share, rendering firms less willing to sacrifice exclusivity by pooling their patents because some discoveries lack substitutes. By increasing the power of some, this lack of substitutes worsens holdout problems.

Because the rights are spread broadly across both public and private institutions that vary widely in their interests and abilities to negotiate, accumulating the required rights is particularly difficult. Moreover, valuation difficulties may prove intractable as the rights cover a large range of technologies, and cognitive biases lead patentees to overvalue their own inventions and undervalue another’s. The net result is that the combination of high transaction costs and strategic behavior may prevent licensing from occurring. The doctrine of blocking patents that was supposed to facilitate licensing simply may not work under such conditions. The cost will be the production of fewer products that benefit human health. Professors Heller and Eisenberg emphasize that to prevent this from occurring, “[p]olicy-makers should seek to ensure coherent boundaries of upstream patents and to minimize restrictive licensing practices that interfere with downstream product development.”

Another solution may be to employ a doctrine of fair use to excuse a certain amount of infringement. For example, the defense of fair use

234. See supra note 230.
235. See supra note 230.
236. See supra note 230.
237. See supra note 230.
238. See supra note 230.
239. See supra note 230.
240. See supra note 230.
241. See supra note 230.
could be used to excuse infringement by researchers attempting to invent around the patent even when the eventual end product is to be marketed commercially.\textsuperscript{242} It might also be used to allow the marketing of an infringing product if the five-factor test supports such a step. For example, if a developer has amassed most of the required licenses to market a product, but a holdout exists, fair use could function as a threat to overcome this hold-up problem. If the remaining patentee refuses to license, a court might allow the infringement to proceed and award a royalty while the infringement continues. Moreover, the mere existence of fair use may make patentees more willing to form institutions to decrease transaction costs.

In fact, the fair use doctrine may be helpful generally in industries characterized by rapid, cumulative innovation where high transaction costs and broad basic patents threaten follow-on innovation. Professors Robert Merges and Richard Nelson have argued that a system characterized by multiple, competing sources of innovation benefits society in the form of more rapid progress than would occur under a regime with few sources of innovation.\textsuperscript{243} They argue that to realize this preferred state, courts should adopt a more nuanced analysis under both the equivalents and reverse equivalents doctrines to help solve patentee hold-up problems.\textsuperscript{244}

An alternative approach would be to use a doctrine of patent fair use to excuse some limited amount of infringement. Unlike the reverse doctrine of equivalents, which leaves the patentee with nothing, the proposed fair use defense at least offers defined factors for courts to con-

\textsuperscript{242} See Eisenberg, Proprietary Rights, supra note 69, at 224–25 (discussing how patents on research tools may impede the progress of science and considering when courts should use the experimental use defense to excuse a certain amount of infringement given the competing interests of various communities).

The case for allowing the defense appears strongest where the subsequent user is attempting to devise alternatives to the patented invention. In such a case, the interests of the research user are congruent with the interests of the public and the scientific community in advancing the state of human knowledge. The patent holder, by contrast, has an interest in prolonging the period in which the public is dependent on the patented technology. If the patentee sees the research user as a competitor rather than a customer, she may refuse to license the invention. Without an experimental use defense, it is possible that no one would be able to build on the inventor’s discovery until the patent expired.

\textsuperscript{243} See Merges & Nelson, supra note 63, at 842–44, 908 (assessing Professor Kitch’s prospect theory of patents and using an array of historical evidence from different industries to question it and argue for the opposite approach). An unstated premise of this Article is that Merges’s and Nelson’s view is correct. Fair use makes little sense if one subscribes to the Kitch model, in which the original patentee is best suited to direct innovation in the field. Under this view, broad patent scope induces more effective development of future inventions.

\textsuperscript{244} See id. at 909–11 (noting that compulsory licensing might also solve hold-up problems but is regarded as anathema).
sider, and can safeguard the original patentee’s incentives by allowing a court to impose a royalty.

C. Fair Use as a Tool to Address Flaws in the Patent System

In addition to its primary benefit of excusing some socially valuable infringement and thereby indirectly helping industries move toward licensing solutions, fair use may also mitigate certain problems created by the Patent Act and offset some of the institutional bias from which the PTO may suffer. More specifically, a fair use defense could help to deal with the “submarine” patents that the Act enables and decrease the cost of invalid patents that at least one study suggests the PTO routinely issues.245

The Patent Act provides certain procedural vehicles that may be used to delay patent issuance. Such patents are referred to as “submarine” patents.246

An inventor . . . files an application with broad claims . . . and then files a series of continuing applications to keep the patent submerged in the patent office; then, one day, someone innocently decides to use the yet to be patented idea, and after they begin production, the inventor surfaces the application through its issuance, and demands the payment of royalties, lest a lawsuit will be filed for infringement.247

While there is some dispute about the number and cost of such patents, they undercut the patent scheme by reducing a competitor’s utility in designing around the patent.248 Such inventing around “is, in fact, one of the ways in which the patent system works to the advantage of the public in promoting progress in the useful arts, its constitutional purpose.”249 The submarine patentee can manipulate the system by amend-

245. See John R. Allison & Mark A. Lemley, Empirical Evidence on the Validity of Litigated Patents, 26 A.I.P.L.A. Q.J. 185, 205–09 (1998) (setting forth results of their study of cases on patent validity that demonstrates that 46% of litigated patents were found invalid, often for failure to meet the non-obviousness requirement); see also infra notes 253–256 and accompanying text (explaining why the PTO erred).


248. See id. (stating that one way to prevent designing around is by “filing a continuation application just prior to the issuance of your patent, and then amending the application’s claims around the competitor’s design changes”). Blount also notes that the impact of submarine patents is in dispute, but that they are likely to increase in light of a particular inventor’s (Jerome Lemelson) success in employing them. See id. at 14–15 (gathering statistics on submarine patents); id. at 19–24 (describing Mr. Lemelson’s tactics).

249. Id. at 12 (quoting Slimfold Mfg. Co. v. Kinkead Indus., 932 F.2d 1453, 1457 (Fed. Cir. 1991)).
ing the application once another company has achieved commercial success, "wreaking havoc on well established industries . . . [and leaving them] scrambling for a defense because the long pendency of the applications permits the applicants to write claims that directly cover industry-standard products."250

While commentators have proposed various solutions to this problem, and Congress has taken some steps that will lessen the number of submarine patents,251 fair use could also help police them. As an equitable defense, it could protect the innocent infringer when the patentee has used the literal terms of the Patent Act to frustrate the system's goals.

While submarine patents result from "gaming" the statute itself, other problematic patents are a product of the PTO's institutional structure. The PTO routinely grants patents on inventions that do not meet the statutory standards.252 These errors result from a number of factors, including inadequate prior art databases,253 limited reexamination procedures,254 and examiner incentive schemes that award compensation for issuing patents, regardless of their validity.255 Lack of information about

250. Dewitt, supra note 246, at 601.

251. See id. at 609–20 (describing how old Supreme Court precedent could be employed in this endeavor); see also Blount, supra note 247, at 27–29 (identifying three proposed ameliorating doctrines: (i) the recent change in patent term from date of issuance plus 17 years to date of filing plus 20 years which provides less incentive to delay issuance; (ii) using the doctrine of laches; and (iii) a proposed amendment to require the publication of applications 18 months after filing). In 1999, Congress passed a modified version of the third alternative by providing for publication 18 months after filing when the applicant has filed in a foreign country or through the Patent Cooperation Treaty. See the Intellectual Property and Communications Omnibus Reform Act, Pub. L. No. 106-135, § 122, 35 Stat. 1501A-521, 1501A-561 (1999).

252. See supra note 245 (describing study documenting percentage of patents incorrectly granted).

253. See, e.g., Julie E. Cohen, Reverse Engineering and the Rise of Electronic Vigilantism: Intellectual Property Implications of "Lock-Out" Programs, 68 S. Cal. L. Rev. 1091, 1175–81 (1995) (outlining defects in PTO handling of software patents); Kobak, supra note 227 at ¶ 5 (noting concern that overly broad patents "may too often be granted by a beleaguered PTO sometimes forced to act on the basis of inadequate disclosures of prior art by applicants," and stating that these patents may discourage further research by creating anxiety for competitors).

254. See Merges, Impossible Patents, supra note 282, at 610–14 (noting that reexamination under American law has been less effective than under European law, and suggesting ways to improve the information available to patent examiners). Under U.S. law, reexamination is limited to certain grounds and will be initiated only if the reexamination request raises "a substantial new question of patentability." 35 U.S.C. § 303 (1994).

255. See Merges, Impossible Patents, supra note 282, at 607 (noting that among PTO employees, bonuses are determined by the number of dispositions and, because of procedural limits, "the only way to earn bonus points with confidence is to allow a patent application").
the prior art and limited reexamination create difficulties, particularly for newly patentable subject matter like software and business methods. 256

Certainly, one way to correct a number of these problems is to improve the internal workings of the PTO. However, chances for reform are rather remote given its entrenched institutional culture. Even if Congress had the will to overhaul the PTO, it is neither humanly nor financially possible to ensure that it will issue only valid patents. A patent fair use doctrine might indirectly decrease the costs of the invalid patents the PTO grants.

While most patents are not litigated, some are licensed. By giving the licensee a bargaining chip, fair use may effectively decrease the royalty associated with the license of an invalid patent, mitigating the distributional and allocative inefficiencies associated with the grant of such a patent.

When patents are litigated, the plaintiff patentee is often subjectively sympathetic and is always objectively entitled to having its patent presumed valid. To the extent that a jury or judge is reluctant to enter an all-or-nothing holding of invalidity, fair use offers another option. It provides for a judgment that the patent is valid and infringed but that infringement is excused, sometimes for free. 257

III. OBJECTIONS AND ALTERNATIVES

In making the case for patent fair use, the preceding Part addressed the issue of whether patent fair use penalizes success (particularly in network markets). However, it did not tackle other objections that policymakers might raise against incorporating a fair use doctrine into patent law. This Part does so, arguing that patent fair use can address a number of situations cost-effectively, may have a beneficial effect on licensing, and is particularly well-suited to be a complement to, and at times a substitute for, antitrust enforcement.

256. See id. at 589 (noting the "disturbing" lack of prior art references in software and particularly in software business methods patent applications).

257. Of course, a court could hold the patent valid and infringed but set damages at zero, effectively arriving at the same result as holding the patent invalid. This would be similar to a fair use finding, although fair use would allow the infringement to continue. However, the patent statute requires that "in no event [shall the court award] less than a reasonable royalty for the use made of the invention by the infringer." 35 U.S.C. § 284 (1994). Under particular circumstances, a court might label a "reasonable royalty" to be zero, but that seems unlikely. See also Grossman, supra note 15, at 262 (noting that the relatively strong damages remedies available under copyright law may help to justify its fair use doctrine, but arguing that it may be advisable to allow more exceptions to patent infringement to develop even though patent law's remedies are less severe than copyright's).
A. Overbreadth and Unintended Results

Fair use might be criticized as an overbroad solution to a limited problem. A tailored response like a statutory compulsory license could address unique issues presented by software API patents or problems raised by network markets more generally. Changes to the experimental use doctrine could partially free researchers from fears of infringement, while the reverse doctrine of equivalents could be clarified to more easily release the radical improver from the strictures of a broad basic patent. To the extent fair use is a response to problems like submarine patents or perverse incentives in the PTO, policymakers could adjust both the Patent Act and the PTO’s internal workings to increase the probability that the system will grant the appropriate set of rights. Why implement a broad fair use defense that should never be successful in many industries but whose existence will complicate dealings universally?

At least two reasons support such a step. First, a case-by-case adjustment of doctrine, either through statutory changes or common law adjudication, is likely to be more expensive than a fair use doctrine that would be enacted only once. Whether changes occur incrementally or through the addition of fair use, costs will increase for a time as parties and courts struggle to understand the new doctrine. A fair use approach will almost certainly prove cheaper than a number of discrete changes because courts will have copyright precedent from which to draw by analogy.

Simply put, fair use has the flexibility to tailor rights across a broad spectrum of industries. Correcting quirks in the patent system and

258. Professors Cohen and Lemley have suggested that patent law codify a right to reverse engineer to achieve compatibility. See Cohen & Lemley, supra note 207, European and American copyright law and domestic law already contain such provisions. See Council Directive 91/250, art. 6, 1991 O.J. (L 126) 42, 45; supra notes 185–185 and accompanying text (discussing how the DMCA adopts the Soya rule); see also The Semiconductor Chip Protection Act of 1984, 17 U.S.C. § 906(a)(1) (1994) (permitting "a person" to reproduce a protected mask work "solely for the purpose of teaching, analyzing, or evaluating the concepts or techniques embodied in the mask work . . ."). Drafters could broaden the exception beyond software to permit reverse engineering generally in the course of producing a non-infringing commercially marketed end product. This would answer the concerns of commentators who fear that patents on basic research tools actually retard further innovation. In a seemingly broader vein, Professor Merges proposes a rule of "technological genericide" that would allow patent law to divest the patentee of its property right if it invents something that becomes a successful standard. See Merges, Charles River Bridge, supra note 77, at 35. At the same time, Merges cautions against unquestioning application of such a principle, emphasizing that the law of takings may have something to say about limiting property rights in this way. See id. at 46.

259. See supra Part I.D.1 (explaining Professor Eisenberg’s suggested changes to the experimental use defense); see also Grossman, supra note 15, at 283 (addressing the experimental use exception in the context of university research).

260. See supra Part II.B (summarizing Professors Merges’s and Nelson’s arguments in favor of an expanded reverse doctrine of equivalents).

261. See supra Part II.C (describing submarine patents and how the inner workings of the PTO encourage issuance of patents rather than issuance of valid patents).
counteracting PTO errors are benefits that are incidental to fair use’s main purpose of providing a unifying yet flexible concept under which courts may address a range of market failures that present recurring problems in a number of distinct contexts.

Second, the basic insight of this Article is that all of the piecemeal changes institutions might make to improve the patent system arise from a common core of issues raised by market defects that share similar characteristics. The anti-patent motives that can frustrate competition in a network market may also lead patentees to refuse licenses to researchers seeking to verify functionality. The externalities that characterize network markets are replicated in the research context and also when the radical improver cannot capture the intangible benefits to society of advancing progress. High transaction costs, strategic bargaining, and bargaining breakdown are universal phenomena not restricted to particular technologies or the non-commercial or commercial contexts. While none of these problems are new to the patent system, they are increasing in the “brave new” patent world and now require an integrated approach. Rather than being overbroad, fair use is a flexible approach to a common problem.

Another more theoretical objection might be directed at fair use. Courts may be more receptive to claims of patent fair use than under copyright law because fair use in patent allows them to impose a fee. This may unintentionally start the system down a slippery slope that transforms the traditional patent property right to exclude into a liability rule where all uses are fair, and the only issue is setting the appropriate royalty.

At its heart, this objection goes to judicial competence in applying the new doctrine. Applied correctly, patent fair use excuses only those infringements that pass the five-factor test. The ability to award a fee is an insurance policy against unduly reducing the patentee’s incentives. It is not an invitation to the judiciary to reform patent law into a liability regime.

That the judiciary will occasionally err and that patentees and the public will bear the costs of those errors is inevitable. However, this is simply not a condition unique to fair use nor a sufficient reason to oppose its implementation. In fact, because the Federal Circuit is generally the exclusive forum for hearing patent appeals, the chances for incorrect precedent are somewhat lower for patent fair use than under the already existing copyright scheme.262

262 See, e.g., William W. Fisher III, Reconstructing the Fair Use Doctrine, 101 Harv. L. Rev. 1659, 1720 (1988) (noting that if authors cannot reliably know what constitutes fair use, their expected income from exploitation of the work will be less certain, causing underinvestment in new works). In other words, if fair use is unpredictable, it will thwart the very goals it is trying to achieve.
B. Impact on Licensing Negotiations and Litigation

But if the goal of fair use is, at least in part, to correct market defects that prevent licensing from occurring, it is counter-intuitive, flying in the face of the conventional wisdom that strong property rights best facilitate licensing. Indeed, it would be a perverse result if, by complicating valuation questions, fair use retarded both private solutions (whether one-to-one licensing, pooling, or standard setting) and the very innovation it was intended to foster. However, while fair use might frustrate private solutions for a time, these effects as well as litigation costs will decline as courts establish a body of precedent. In particular, fair use can play a useful role in disciplining patent pools and eventually encouraging private standard-setting.


a. Patent Pooling. — Although commentators applaud private institutional arrangements to facilitate licensing, there are actually few patent pools in existence today, perhaps in part because of their vulnerability under antitrust law. Generally, pooling is most likely to occur when there are multiple rightholders, valuation is relatively simple, the members of the pool engage in repeat transactions, and joining the pool lowers the cost of excluding others by going it alone. At least at first glance, the argument that strong property rights are required for a patent pool to form is persuasive. Accordingly, the weaker patent right proposed in this Article might appear harmful to private licensing and pooling efforts.

However, Professor Merges has found that under certain conditions, weak rights may not frustrate private orderings. In particular, where there is a “close-knit group of experts with shared understandings of the technology, industry, and entitlements structure” that is extracting monopoly rents and successfully enforcing its own private code of conduct,

263. See, e.g., Ashish Arora, Licensing Tacit Knowledge: Intellectual Property Rights and the Market for Know-How, 4 Econ. Innov. New Tech. 41, 42 (1995) (arguing that “broader patents improve the efficiency of technology transfer”); Dam, supra note 27, at 267 (“[I]t is important that the line between the patented and the unpatented be clearly demarcated in the patent itself, rather than being left to future litigation, so that a green light is given to R&D beyond that line”); Lemley, Improvement, supra note 10, at 1055 (“Uncertainty over the value of an invention ... can prevent the parties from agreeing to an efficient license transaction. So can uncertainty over the scope of a patent—that is, over whether the licensing transaction is necessary at all.”); cf. Robert P. Merges, Intellectual Property Rights, Input Markets, and the Value of Intangible Assets 49 (1999) (unpublished draft on file with the Columbia Law Review) (noting that clearer intellectual property rights may facilitate licensing, particularly in input markets, but cautioning that “effects to define ownership rights will not always increase efficiency”).

264. See Heller & Eisenberg, supra note 230, at 709 (“Although antitrust law may be less hostile to patent pools today than it was in 1975 when a consent decree dismantled the aircraft patent pool, the antitrust climate changes from one administration to the next. Even a remote prospect of facing treble damages and an injunction may give firms pause about entering into such agreements.”).

265. See Merges, Liability Rules, supra note 82, at 1301–25.
weak property rights will not prevent informal institutions from arising. 266

In many industries, however, including software and biotechnology, while some of the players might be "repeat," there is rapid market turnover, making the existence of a "close-knit group of experts" somewhat unlikely. The question is whether fair use would so complicate valuation decisions that rightholders would fail to form a private institution to facilitate licensing. It may, at least for a time. However, as precedent decreases the cost of fair use by making its contours more certain, so too would it lessen valuation difficulties.

Additionally, fair use might perform a useful function in policing patent pools. The patent collective may fail to consider the interests of future inventors, thus imposing negative externalities on the public. 267 Moreover, there is a reason why antitrust law is skeptical of such arrangements: They may mask the operation of a cartel. Fair use would make the pool more responsive to competition if new market entrants could use it as a credible threat.

b. Standard-Setting. — Closely allied to the concept of patent pools is standard-setting, where companies or other parties come together to adopt a compatibility regime voluntarily. 268 Standards bodies like the American National Standards Institute (ANSI) may allow proprietary technology to be included in a standard specification. 269 However, ANSI Guidelines encourage early disclosure of patents and require the patentee to agree to make the technology available on reasonable terms. 270

The prevalence of standards bodies alleviates some pressure on the patent system and explains why lobbying for a compatibility principle or fair use doctrine in patent law has been slow to develop. But not all in-

266. Id. at 1369 (summarizing the similarities between a medieval guild and the Fashion Originator's Guild of the 1930s).


268. [S]tandards in the United States have traditionally been developed by various types of groups, including [government,] trade associations, professional societies, general membership organizations, and third party certifiers (currently, some 400 in all). Those standards which are set by official organizations, such as the American National Standards Institute (ANSI), are referred to as de jure standards, and are often granted the greatest recognition . . . . At the opposite end of the spectrum are those product standards, no less influential, which have been established by sheer market power or other dynamic forces by one or a very small number of commercial vendors . . . . In the middle lie those other variously formal and informal (and quite numerous) alliances of companies which set . . . de facto technical standards.


270. See id.
dustries standardize. The computer industry, while it has a large number of standards organizations, has been driven by the de facto market-created proprietary standards of Intel (hardware) and Microsoft (software) that are not encumbered by an obligation to make them available on reasonable terms.

In fast-moving, high-tech markets generally, standardization may be less feasible than in conventional markets, because standards organizations are notoriously slow to act, and subject to capture by a dominant player. Generally, "a vendor will rationally favor standards only if the value of its own product is enhanced at least as much as is the value of its rival's. Often, a small firm's products benefit more from standards than do a large firm's." 271

By weakening the patent right, fair use decreases the expected value to firms in a network market of going it alone, making it more likely that they will voluntarily set open standards. While there are both pros and cons to standardizing technologies, 272 as long as companies remain free to implement the standard competitively and innovate around it, consumers should benefit from expanded choice and lower prices.

2. Litigation Costs. — The same uncertainty that could make licensing less likely under a patent fair use regime will increase litigation costs by complicating both matters of proof and the calculation of an appropriate remedy. But while the first few cases raising the defense are likely to be expensive, costs will decline as courts establish precedent that enhances predictability. Again, the Federal Circuit's existence will be helpful in minimizing errors and inconsistencies.

Further, the only solution for the system as currently constituted is for courts to develop the appropriate expertise to deal with complex technologies and new market realities. 273 While commentators (rightly or wrongly) usually view the legislature as comparatively better-suited to gathering data and formulating judgments, it lacks the relevant information at the appropriate time to fine-tune intellectual property rights, leaving the courts as the institution of last resort. The courts, in addition to informing their decisions with reference to copyright cases, may also need to transfer skills that they have developed in adjudicating other issues. For example, antitrust cases already involve complicated economic evidence. Courts could use the expertise developed there in applying the


272. See supra Part II.A.1 (noting that a standard may be both inferior and durable, leading to a sub-optimal state).

273. Moreover, Congress seems willing to impose such burdens on them. For example, in at least one of the database protection bills (the Cobie bill) that Congress is considering, the drafters have adopted a five-factor test reminiscent of copyright fair use that focuses on market harm in considering whether to excuse a person's unauthorized extraction and use of information contained in a protected database. See H.R. 354, 106th Cong., 1st Sess. at § 1403. See also supra note 121 (discussing the Bilby bill, another proposed database protection regime).
five-factor test both to decide whether to label the infringement "fair" and whether to charge for it.

3. The Benefits of Uncertainty. — While fair use will inevitably introduce uncertainty into the patent system, at least for some period of time, recent work demonstrates that the assumption that uncertainty is unequivocally a "bad thing" is incorrect.274 Because the last bit of monopoly pricing by a patentee with market power imposes large social costs while generating only small benefits for the patentee, judiciously constraining that power may not harm incentives to innovate.275 One way to limit such power is by introducing some uncertainty into the scope of the patent grant.276

Some commentators suggest a system of probabilistic patents in which a patentee has a right to seek damages only after the patent's expiration, and courts award those damages with some probability less than one.277 The uncertainty and delay built into such a system would induce a limited amount of infringement that would not decrease the ex ante innovation incentive, particularly if coupled with an increase in patent length.278

In contrast to a system of probabilistic patents, fair use would offer a more targeted way to introduce uncertainty when appropriate.279 Be-

---

274. See Ayres & Klemperer, supra note 5, at 988 (arguing that a system with some uncertainty and delay can more efficiently provide the appropriate reward to the patentee).

275. See id. at 987; see also id. at 989–93 (graphically illustrating how the last increment by which a patentee increases price imposes costs on society that exceed the benefit to the patentee). Ayres and Klemperer also note that in network markets, patentees with complete monopoly power may charge less than the static monopoly price but that even in such cases, their insights still hold. See id. at 1016–17.

276. Recent research reveals that the "possibility that valid patents will be unenforced (Type I uncertainty) is more likely to be efficient than the possibility that invalid patents will be enforced (Type II uncertainty)." Id. at 1019–20. Underinclusive standards are more likely than rules, or overinclusive standards, to create the uncertainty and delay that would induce limited infringement and create the more efficient Type I uncertainty. See id. Patent fair use is likely to function as an underinclusive standard, increasing the risk that valid patents will be unenforced or at least under-enforced. Patent fair use then, by introducing uncertainty, may increase the efficiency of the system under Ayres' and Klemperer's view.

277. See id. at 993–1000.

278. See id. at 993, 1001–07 (explaining how limited infringement would constrain the patentee's market power and why expanded duration could maintain the incentive to innovate).

279. While Ayres and Klemperer do not advocate "that courts flip coins after fifty years to decide whether patents should be enforced," and do make certain concrete proposals, see id. at 1019–52 (offering ways to encourage limited amounts of infringement, including making it more difficult for patentees to obtain preliminary injunctions, placing decisionmaking with less specialized entities, and offering alternatives to increasing patent duration including expanding scope and awarding partial damages), their primary focus is on identifying uncertainty and delay as illuminating policy questions. Patent fair use may induce less infringement than Ayres and Klemperer would view as desirable but at least would be specifically tied to the system’s goals.
cause it is informed by reference to defined factors geared toward realizing the goals of the patent system, it is likely to be quite sensitive to fine-tuning the rights in particular situations. It would not introduce systemic uncertainty, particularly once patentees become aware of the paradigmatic cases in which it will likely be successful. Thus, fair use would tailor the level of uncertainty to avoid a perverse result yet help to achieve the socially optimal level of infringement.

C. Patent Within the Larger Legal Context

This section began by noting that modifications of existing doctrines and to the workings of the PTO could be combined to create the equivalent of fair use in the field of patent law. While this is not the preferable approach for the reasons discussed above, it does highlight an important point: To the extent that policymakers change one or more pieces of the legal system, the need to adjust others may correspondingly increase or decrease.

Moreover, the relevant “system” includes more than simply the intellectual property statutes. Intellectual property law interacts with other public and private law that can either assist or frustrate it in achieving its goals. In particular, antitrust law has cast a long shadow over intellectual property law, constraining rightholders from engaging in anti-competitive conduct. One question is whether patent fair use is needed in addition to antitrust law and, if so, whether antitrust law should be more accommodating of patent pools and standard setting efforts to reinforce the beneficial effects of fair use.

While antitrust law shapes an intellectual property rightholder’s expectations with regard to how it may exercise that right, patent fair use is still necessary. Patent fair use would address situations in which the antitrust laws may not be implicated and therefore have no impact on the patentee’s decision to license. For example, there may be no threat of the patentee’s obtaining an economic monopoly in many research and educational contexts, but it still might make sense to excuse some infringement as fair. In commercial contexts too, where network effects are strong, the tipping point at which policymakers might want to open a patented interface to competition may fall short of the market share required to sustain an antitrust monopolization claim. Moreover, in non-network markets, where defects tend to be high transaction costs resulting from a splintering of rights, antitrust concerns are unlikely to be implicated until rightholders choose to pool their patents. Finally, fair use, with its appropriate sensitivity to the balance that patent law seeks to achieve, is better suited to addressing the propriety of unilateral refusals.

---

280. Reichman, supra note 34, at 2456–37 ("A reevaluation of intellectual property rights must recognize the interplay of those other disciplines in different economic environments in order to represent accurately the balance between legal incentives to create and free competition.").
to license than is antitrust law (as that law's tortured treatment of such questions attests). 281

Antitrust law, then, is not a substitute for patent fair use, but rather a supplement to it. It is a part of the public law that should be sensitive to the pro-competitive benefits that private licensing institutions, including patent pools, may offer, and the role of fair use in encouraging these solutions and safeguarding against their abuse.

Leaving patentee conduct to policing by antitrust law may not be the best systemic approach. If patent law does not clearly adopt rules allowing infringement where appropriate, antitrust law may step in to correct patent law's errors with its sledgehammer-like remedies including treble damages and structural relief. Moreover, by the time antitrust law is pressed into action, a court may have few remedies available to it that will effectively redress the accrued antitrust injury. 282 Some, like structural relief, are likely to be quite drastic. Fair use, if implemented correctly, should intervene in the market at a time when it is still possible to maintain competition. Ideally, fair use should decrease the need for later invasive and expensive antitrust actions. As the economy becomes increasingly complex, courts will have to develop skills in addressing market conditions, whether in antitrust or intellectual property cases. From a social perspective, these skills could probably be most effectively deployed in intellectual property disputes rather than in antitrust litigation.

CONCLUSION

That the intellectual property system is imperfect has obvious and important consequences for the society that bears the costs of that imperfection. However, constructing a system that never errs would entail costs beyond those that society is—or should be—willing to bear. But it does not logically follow that there are no cost-effective improvements that would help increase the probability that the system will grant socially optimal rights.

This Article argues that patent law should adopt a fair use doctrine to help it prevent rights from becoming overbroad in the new circumstances of today's high-tech world. The defense would authorize courts

281. See Maureen A. O'Rourke, Striking a Delicate Balance: Intellectual Property, Antitrust, Contract, and Standardization in the Computer Industry, 12 Harv. J.L. & Tech. 1, 25–31 (1998) (detailing antitrust precedent addressing unilateral refusals to license and noting the conflicting precedent and the questions it raises); see also In re Indep. Serv. Org. Antitrust Litig., 203 F.3d 1322 (Fed. Cir. 2000) (distancing itself from other courts and holding that a patentee's refusal to license is not subject to antitrust scrutiny unless the patent was obtained fraudulently or the infringement suit is "both objectively baseless and subjectively motivated by a desire to impose collateral, anti-competitive injury rather than to obtain a justifiable legal remedy").

to weigh defined factors in deciding whether or not to excuse an infringement as fair. The factors themselves are designed to help courts identify and address market failures that may prevent socially efficient innovation from occurring. While modeled after its copyright counterpart, unlike that doctrine, the defense in patent law would allow a court to consider the same factors in deciding whether a fair use should be excused without compensation from the infringer to the patentee. This approach guards against inadvertently reducing patentees' incentives through invocation of the fair use doctrine. Moreover, because the defense is broadly applicable to market failures across a spectrum of industries, its costs of implementation should, at least as its application becomes more predictable, be offset by its benefits.

Moreover, as the economy becomes increasingly global, pressure to harmonize intellectual property law is increasing. Because countries vary in their state of development and policy interests, a rigid, uniform system is undesirable. A flexible doctrine of fair use that could be universally adopted will, by its very nature, allow nations to retain some autonomy in interpreting intellectual property law in a principled way to accord with local values.

Finally, while some might object to the introduction of a new patent doctrine that weakens the patent right, failure to act is not a viable option. Throughout history, every suggested patent reform that would weaken exclusive rights has been met with at least some dire predictions of doom for the patent system and innovation. But rights that are too strong can be equally costly, by denying society the benefit of follow-on innovation. Fair use is a flexible doctrine that safeguards the incentive to invent, while providing a safety valve to ensure that the patent system does not perversely function to impede the very innovation it is intended to encourage.