A PROPERTY RIGHTS THEORY OF THE LIMITS OF COPYRIGHT

by

Norman Siebrasse

Assistant Professor, Faculty of Law, University of New Brunswick

siebrass@unb.ca
Faculty of Law
University of New Brunswick
Fredericton, N.B.
Canada, E3B 5A3

[W]e must take care to guard against two extremes equally prejudicial; the one, that men of ability, who have employed their time for the service of the community, may not be deprived of their just merits, and the reward of their ingenuity and labour; the other, that the world may not be deprived of improvements, nor the progress of the arts be retarded.[1]

"[C]opyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work."[2]

"The various doctrines of copyright law, such as the distinction between idea and expression ...can be understood as attempts to promote economic efficiency by balancing the effect of greater copyright protection – in encouraging the creation of new works by reducing copying – against the effect of less protection – in encouraging the creation of new works by reducing the cost of creating them."[3]

INTRODUCTION

From the Lord Mansfield, to the Supreme Court to Landes and Posner, the view that copyright law should seek to strike a balance between providing an incentive to create works and encouraging their dissemination is firmly entrenched in judicial decisions and academic commentary.[4] The intuition is straightforward. Without copyright protection, authors of new

---

[1] *Sayre v. Moore* (1785), quoted in *Cary v. Longman*, 1 East *358, 362 n. (b), 102 Eng. Rep. 138, 140 n. (b) (1801).

[2] *Feist Publications Inc v Rural Telephone Service Co. Inc* 111 S. Ct. 1282, 1290 (1991).

[3] William M. Landes & Richard A. Posner, An Economic Analysis of Copyright Law, 18 J. Legal Stud. 325, 333 (1989).

[4] See also *Twentieth Century Music Corp. v. Aiken* 422 U.S. 151, 156, 95 S. Ct. 2040, 2043-4 (1975), "The limited scope of the copyright holder's statutory monopoly. . . reflects a balance of competing claims upon the public interest: Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts."; *Harper & Row v Nation Enterprises* 723 F.2d 195, 206 (1983) "The Copyright Act provides two mechanisms by which the rights of a copyholder may be protected without impeding the public's access to information. The first of these devices, the distinction between expression, which is copyrightable, and idea or fact, which is not, has been discussed above. The second means of ensuring a proper balance of the citizenry's need to be informed and the author's monopoly of his original writings is known as the doctrine of "fair use."; *Whelan Assoc. Inc. v Jaslow Dental Laboratory Inc*., 797 F.2d. 1222, 1235 (3rd Cir. 1986), "...we must remember that the purpose of the copyright law is to create the most efficient and productive balance between protection (incentive) and dissemination of information, to promote learning, culture and development."; *Computer Assoc. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 696 (2d Cir. 1992), "the copyright law seeks to establish a delicate equilibrium. On the one hand, it affords protection to authors as an incentive to create, and, on the other, it must appropriately limit the extent of that protection so as to avoid the effects of monopolistic stagnation.."; *Gates Rubber Co. v. Bando Chem. Indus., Ltd.*, 9 F.3d 823, 839 (10th Cir. 1993), "Copyright policy is meant to balance protection, which seeks to ensure a fair return to authors and inventors and thereby to establish incentives for development, with dissemination, which seeks to foster learning, progress and development.";

works will be undercut by pirates who do not have to bear the costs of creation. Anticipating this, authors will not enter the market in the first pace, and too few works will be created. But on the other hand, progress is cumulative, and with too much protection authors will not be able to afford to use the prior works which they build on in creating new works. Again, too few works will be created. Somewhere between these two extremes of protection lies a happy medium. So, we conclude, the aim of the law should be to find this ideal level of protection.

This balancing story is persuasive, and with good reason. It is correct – except in the conclusion. While there is indeed a level of copyright protection which strikes the optimal balance between incentives to create and restrictions on dissemination, it does not follow that the goal of the law should be to strike that balance. Determining the optimal level of protection requires far more information than is available to the courts, or, for the most part, to the legislature, and the attempt to achieve the impossible goal of striking the optimal balance is likely to do more harm than good.

This is essentially the argument made by Coase against Pigovian taxes as a means of correcting externalities. Both copyright and environmental law aim to protect the interests of

---

Hustler Magazine, Inc., v Moral Majority, Inc., 796 F.2d 1148, 1151 (9th Cir. 1986) "The fair use doctrine is "is a means of balancing the need to provide individuals with sufficient incentives to create public works with the public's interest in the dissemination of information."; *Cardtoons, L.C., v. Major League Baseball Players Association*, 95 F.3d 959, 976 (10th Cir. 1996) "One of the primary goals of intellectual property law is to maximize creative expression. The law attempts to achieve this goal by striking a proper balance between the right of a creator to the fruits of his labor and the right of future creators to free expression. Underprotection of intellectual property reduces the incentive to create; overprotection creates a monopoly over the raw material of creative expression."; *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 974 (1980): "The copyright provides a financial incentive to those who would add to the corpus of existing knowledge by creating original works. Nevertheless, the protection afforded the copyright holder has never extended to history, be it documented fact or explanatory hypothesis. The rationale for this doctrine is that the cause of knowledge is best served when history is the common property of all, and each generation remains free to draw upon the discoveries and insights of the past."; and see also *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984) (fair use), *Herbert Rosenthal Jewelry Corp. v. Kalpakian* 446 F.2d 738, (9th Cir. 1971) cited *infra* n.000 and *Computer Assoc. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 696 (2d Cir. 1992) cited *infra* n.000. From Congress: "In enacting a copyright law Congress must consider . . . two questions: First, how much will the legislation stimulate the producer and so benefit the public; and, second, how much will the monopoly granted be detrimental to the public? The granting of such exclusive rights, under the proper terms and conditions, confers a benefit upon the public that outweighs the evils of the temporary monopoly." H. R. Rep. No. 2222, 60th Cong., 2d Sess., 7 (1909). Among academics writers see Stephen Breyer, The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies and Computer Programs, 84 Harv. L. Rev. 28 (1970); Paul Goldstein, COPYRIGHT: PRINCIPLES, LAW AND PRACTICE (1989) §1.2.2.4; Zechariah Chafee, Jr., Reflections on the Law of Copyright: I, 45 Colum. L. Rev. 502 esp. at 511 (1945); NIMMER ON COPYRIGHT §1.10[B][2]; and of course, Landes & Posner, *id.* For economic treatments, see sources cited *infra* n.000.

consumers (to be free from pollution, to use works freely) while preserving adequate incentives for producers. Pigou sought to strike this balance by directly allocating the right (to pollute, or to be free of pollution) to the highest value user. Coase argued in response that any such attempt, while desirable in principle, was bound to fail in practice because of the information required to strike the correct balance.[5] Like the Pigovian approach to environmental law, the balancing approach to intellectual property advocates an attempt to minimize transaction costs by allocating rights directly to the highest value user. And like a Pigovian tax, balancing is desirable in principle, but is not feasible in practice because it is too informationally demanding.

As a result of Coase's insight, the importance of creating clearly defined property rights in order to prevent market failure is now well recognized.[6] In that spirit, this Article argues that rather than attempting an optimal allocation of rights, the main task of copyright law should be to ensure that property rights are clearly defined. This Article also argues that adopting this goal is consistent with existing copyright law and policy, as many of the key doctrines of copyright law are best explained as being aimed at ensuring clarity in intellectual property rights.

This Article begins with a critique of the balancing approach. Because the problem with the balancing argument is not in principle, but in the execution, it is very persuasive when stated in general terms. A more precise description of the balancing approach, developed in Part I.A, shows how difficult the balancing approach is to actually apply. In particular, the information required to undertake the balancing rigorously is far beyond what is available to the courts. Because the requisite information is simply not available, evolutionary arguments which depends on structural factors, such as litigation pressures, cannot rescue the balancing approach. Part I.B then applies the critique in the specific context of copyright in ideas, arguing that the balancing arguments traditionally made against granting copyright are not supported by empirical information or reliable intuitions.

---

[5] Coase first made the point in The Problem of Social Cost, 3 J. Law & Econ. 1 (1960) at 41-42, but it was apparently insufficiently appreciated, and he later emphasized and elaborated on it in Notes on the Problem of Social Cost, in THE FIRM, THE MARKET AND THE LAW (1988), Part VI and esp. at 184.

[6] An early elaboration of this implication of The Problem of Social Cost is Harold Demsetz, The Exchange and Enforcement of Property Rights, 7 J. Law & Econ. 16 (1964) esp at pp.16-19. See generally Yoram Barzel, ECONOMIC ANALYSIS OF PROPERTY RIGHTS (1989).

Part II offers an alternative property rights approach to copyright doctrine. Part II.A. develops the basic argument in the context of a discussion of the originality requirement in copyright. Part II.B moves on to the related but more difficult issue of copyright in ideas. The specific issues addressed, include how the line between ideas and expression should be defined; why copyright should be denied to ideas, so defined; why some ideas should be protected by patent and not copyright; and how the line between the two regimes should be drawn; and why some ideas should be denied protection entirely. After the core of the theory is developed in Part II, Part III applies the property rights approach to some miscellaneous issues, including copyright in facts, news and history, fair use, and copyright in derivative works. Part IV concludes.

While the property rights approach is an economic analysis of copyright law, it is not premised on a strong efficiency thesis. I claim that the property rights approach explains the thrust of the doctrines in issue, but not every detail. In particular, while the informational requirements of the property rights approach are much less than that of the balancing approach, strictly optimal decisions, even under the property rights approach, would require more information than is available to the courts. Significant uncertainty remains even after the property rights analysis is applied, and the way in which this uncertainty is resolved in specific cases can only be explained by recourse to factors such as prevailing ideology, historical accident, or the views of the individual judge.[7]

Further, I do not claim the property rights analysis explains all aspects of copyright. In particular, the property rights analysis does not explain or justify the term of copyright protection. Since any term of years is equally clear, ensuring clarity in property rights cannot explain why one term is better than another. It may be that structural pressures on term length are sufficiently strong to lead to rough optimality on a balancing analysis, or it may be that ideological or public

[7] Mark J. Roe , Chaos and Evolution in Law and Economics, 109 Harv. L.Rev. 641 (1996) provides persuasive reasons for doubting the strong efficiency thesis. I would further argue that many of the effects the Roe describes as "accidents", and which ultimately have a large impact because of chaos and path dependence, can be explained in terms of the prevailing ideology and judicial psychology. So, I believe that "critical" analyses of copyright, which emphasize the role of the rhetorical or authorship in the development of copyright law (see the authors discussed in Rosemary Coombe, Challenging Paternity: Histories of Copyright, 6 Yale J L & Hum. 397 (1994)) are not necessarily inconsistent with an economic analysis, but rather can be used to complement it.

choice explanations are needed.[8] Also, while some aspects of copyright which I do consider, in particular the law regarding derivative works, can be justified on the a property rights analysis, it seems likely that a political explanation is more descriptively accurate as an explanation for the state of the law.[9]

PART I THE BALANCING APPROACH

I.A Balancing as Allocation of Property Rights

Landes & Posner remark that from an economic perspective, the "distinguishing characteristic" of intellectual property is that it has significant sunk costs of creation and low costs of copying, giving it "public good" characteristics: "While the cost of creating a work subject to copyright protection – for example, a book, movie, song , ballet, lithograph, map, business directory, or computer software program – is often high, the cost of reproducing the work, whether by the creator or by those to whom he has made it available, is often low. If copies made by the creator of the work are priced at or close to marginal cost, others may be discouraged from making copies, but the creator's total revenues may not be sufficient to cover the cost of creating the work."[10] Since free entry into the market will result in marginal cost pricing, the author will realize that she will not be able to recover her sunk costs, and so she will choose not to create the work in the first place.[11]

---

[8] See the discussion *infra* Part I.E.

[9] The right of authors to "dramatize or translate their own works" was granted by the Copyright Act of 1870, ch.230, §86, 16 Stat. 198, 212. The previous Act had been interpreted narrowly, so that translations and abridgements were not infringements, and protection against dramatizations was "thin" – an infringement would be found only if a play based on a novel used large portions of the exact language from the novel: see the discussion in R. Y. Libott, Round the Prickly Pear: the Idea-Expression Fallacy in a Mass Communications World, 14 UCLA L.Rev. 735, 744 (1967); and Goldstein *supra* n.000, §5.3. It seems plausible that the reason that dramatizations where not protected before that time under American law was that authors were largely English, and dramatizers were American. As American novelists came into their own, the pressure grew to protect the rights of authors to dramatize. It is interesting that Landes & Posner, who are famous for their thesis that the common law is efficient, argue that copyright doctrine is efficient even though copyright is primarily statutory: see generally Landes & Posner *supra* n.000.

[10] Landes & Posner, *supra* n.000 at 326.

[11] *Id.* at 328.

Some new works will be created even without copyright rights, because the author will have other advantages, such as a first-mover advantage, lower reproduction costs, or direct sponsorship, which will be sufficient in some cases to allow her to recover her costs of creation.[12] But even so, without the promise of copyright protection the threat of piracy will certainly reduce incentives to create. Intellectual property law addresses this problem by granting the creator of a work a property right allowing her to enjoin at least some forms of copying. The creator can then recover her sunk costs by charging a price above marginal cost.

The difficulty with this solution is that the author's monopoly allows pricing above marginal cost and this results in inefficient restrictions on the dissemination of the work.[13] Copyright protection therefore gives rise to a trade-off between incentives to create a work and restrictions on its dissemination.[14]

It seems to follow naturally from this that copyright doctrine both is and should be about this trade-off. Again, from Landes & Posner, "Striking the correct balance between access and incentives is the central problem in copyright law. For copyright law to promote economic efficiency, its principal legal doctrines must, at least approximately, maximize the benefits from

---

[12] This is emphasized by Breyer, The Uneasy Case for Copyright, *supra* n.000.

[13] Strictly, dissemination is restricted when the price to the marginal consumer is greater than marginal cost. So, if the pricing scheme is such that only infra-marginal consumers pay more than marginal cost, dissemination will not be reduced. This type of pricing scheme will be very rare (see discussion *infra* n.000 and accompanying text), so as a shorthand we can say that pricing above marginal cost restricts dissemination.

[14] The classic statement of this position is that of Macaulay in his speech in the House of Commons on the reading of the Copyright Bill in 1841: "It is desirable that we should have a supply of good books; we cannot have such a supply unless men of letters are liberally remunerated.... Monopoly is an evil... For the sake of the good we must submit to the evil; but the evil ought not to last a day longer than is necessary." See T.B. Macaulay, SPEECHES ON POLITICS AND LITERATURE (1909) 178-180. A more explicitly economic analysis in the same vein is Arnold Plant, The Economic Theory Aspect of Copyright in Books, [1934] Economica 167. In copyright, the most complete general treatment is Landes & Posner, *supra* n.000 ; see also R.M. Hurt & R.M. Schuchman, The Economic Rationale of Copyright, 56 Am.Econ. Rev. 420 (1966). Important economic analyses in the patent context include William Nordhaus, Invention, Growth and Welfare: A Theoretical Treatment of Technological Change, (1969) and F.M. Scherer, Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation, 62 *ARE* 422-27 (1972); and see Plant, The Economic Theory Concerning Patents for Inventions, [1934] Economica 30, emphasizing the diversion of resources into patentable activities as the main drawback of the patent system, rather than restrictions on dissemination of the patented work. For a review see Besen & Raskind, An Introduction to the Law and Economics of Intellectual Property, 5 J. Econ. Perspectives 3 (1991). There is a very large literature which explores more technical aspects of this tradeoff; see the symposium in 21(1) RAND J. Econ., and the literature on optimal patent life, discussed *infra* n.000

creating additional works minus both the losses from limiting access and the costs of administering copyright protection."[15]

This conclusion is too strong. While selective denial of protection may lead to welfare gains, it is equally true that it can lead to welfare losses. If we deny protection in the wrong markets, for example the market for the expression of the original work, the disadvantage of reduced incentives to create would more than outweigh the benefits of increased dissemination of those works which are created. If the balancing analysis is to justify the current limits on copyright protection, it is not enough to show that it is *possible* that denial of protection in some market will lead to net gains. Rather, it must be shown that denial of protection in a specific market is *likely* to lead to gains rather than losses.

This argument parallels Coase's critique of Pigovian taxes as a solution to the problem of externalities. *The Problem of Social Cost* is most famous for the Coase theorem, which states that in the absence of transaction costs rights will be reallocated to maximize the value of production, regardless of the initial allocation of entitlements. But Coase emphasized that since transaction costs are not in fact zero, rights may not be reallocated optimally through the market, in which case the initial allocation of rights will affect the ultimate use of resources.[16] In principle, then, a tax may be levied, for example on a polluter, which, by reflecting the amount which those harmed would be willing to pay to avoid the harm, will improve the final allocation of resources. The difficulty is that calculating the appropriate tax would require knowledge of the net harm caused at different levels of production to all parties affected by the pollution. This is an enormous informational burden, and Coase concluded that "there is, as I see it, no way in which the information required for the Pigovian tax scheme could be collected.[17]

From the Coasian perspective, prohibition of a harmful activity is an attempt to avoid

---

[15] Landes & Posner, *id.* at 326.

[16] See The Problem of Social Cost *supra* n.000, Part VI and Notes on the Problem of Social Cost *supra* n.000, Part VI. Coase has remarked (Notes on the Problem of Social Cost at 174) that "the world of zero transaction costs has often been described as a Coasian world. Nothing could be further from the truth. It is the world of modern economic theory, one which I was hoping to persuade economists to leave."

[17] Notes on the Problem of Social Cost *supra* n.000 at 184. And more picturesquely, "My point is simply that such tax proposals are the stuff that dreams are made of. In my youth it was said that what was too silly to be said may be sung. In modern economics it may be put into mathematics." *id.* at 185.

transaction costs by allocating the right in question (to pollute, or to be free from pollution) directly to the highest value user. This is clearly desirable if it can be done; it is necessary to the optimal allocation of resources if transaction costs are so high that private reallocation of the right is impossible, and an optimal initial allocation of rights will at least avoid the costs of private negotiation. The difficulty is that, as a practical matter, it can't be done. While Coase did not address the issue of directly prohibiting harmful conduct in detail, the argument is the same, since we need at least as much information regarding the overall harm caused in order to determine whether the activity should be prohibited as we do to determine the optimal tax.[18]

Similarly, the balancing approach to copyright is fundamentally an attempt to reduce transaction costs involved in the reallocation of rights by allocating those rights directly to the highest value holder. Because intellectual property is non-exclusive in consumption, allocating the good to the user rather than the creator is achieved by denying the property right to the author rather than by granting a property right to the user. The obvious benefit is that if the right is initially allocated to the author, transaction costs may prevent the user from bargaining for the right, even though she could benefit. But the disadvantage is that denying the property right to the author reduces incentives to create new works. The benefits of free use versus the disincentives to production are the same factors which a decision to prohibit polluting activity must balance. And as in the case of pollution discussed by Coase, the problem with this strategy is that in order to carry it out we need to determine, with some reasonable confidence, who is the highest value holder of the right.

This problem is crucial, because, as Coase noted, unless the Pigovian tax is reasonably accurate, the distortion caused by the inaccurate tax is likely to be just as harmful as the distortion caused by no tax at all.[19] Similarly, an erroneous allocation of copyright will do more harm than good, since denying protection reduces incentives to create at the same time as it

---

[18] In fact, direct prohibition of an activity requires even more information that a Pigovian tax. The Pigovian tax would take the form of a pollution/liability schedule which would be presented to the polluter, who would then choose his level of production: see Notes on the Problem of Social Cost at 183. Thus the government need not know the polluter's cost schedule. If the level of activity is to be directly regulated, the government needs information regarding the polluter's preferences as well as those of harmed by the pollution.

[19] The Problem of Social Cost *supra* n.000 Part VIII.

reduces transaction costs. So, while there is no objection to the balancing approach in principle, the objection is that in practice the necessary analysis is too complex to arrive at firm conclusions through intuition alone, and we do not have the information necessary to support the balancing argument on a detailed quantitative basis.[20]

Coase considered the informational obstacles to a Pigovian tax be so obvious that he initially felt there was no need to spell them out in detail.[21] Given the general acceptance of the balancing approach to intellectual property, such a cursory dismissal will not suffice.[22] Accordingly, the next section of this Article considers the informational requirements of the balancing analysis in general terms. The following section applies the general critique to the specific issue of copyright in ideas by asking whether the traditional arguments against copyright protection for ideas are persuasive in light of these informational requirements.

---

[20] This also echoes Richard Posner's comments on "the inability of the law to measure preferences accurately," and his suggestion that as a result "it is desirable, so far as is consistent with achieving efficient use of resources, to minimize the necessity for broad cost-benefit analysis in legal decisions" ECONOMIC ANALYSIS OF LAW, 2ND ED (1977) at 403. Despite this observation, a broad cost-benefit analysis is precisely what is required by the analysis of copyright law offered in his article with Landes *supra* n.000. (This discussion has vanished by the 4th edition. It is not clear whether this marks a shift in Posner's thinking.)

[21] See *supra* n.000.

[22] William W. Fisher III, Reconstructing the Fair Use Doctrine, 101 Harv. L. Rev. 1659 (1988) esp. Part IV, offers a critique of the balancing approach in the context of the fair use doctrine which similar in spirit to that offered in this Article, although it aims at the individualized balancing approach (see *infra* Part I.D), as is appropriate in the context of fair use. Fisher pursues two avenues in "reconstructing" fair use. While he acknowledges the difficulty in principle of conducting the balancing analysis, he argues that the approach remains useful, and that the doctrine could and should be changed in ways which would bring it closer to implementing an approximate balancing (see esp. Part IV.E). As I discuss *infra* Part III.B, the fair use doctrine does have a necessary balancing component, and Fisher's analysis in this respect is generally persuasive. Fisher also argues that the doctrine should be used to "advance a substantive conception of a just and attractive intellectual culture" (at 1744). As noted *supra* n.000, I believe that non-economic arguments can in principle supplement an economic analysis, in the manner that Fisher suggests, but assessing the merits of Fisher's particular approach is beyond the scope of this Article. James Boyle, in Ch.4 INFORMATION ECONOMICS, OF SHAMANS, SOFTWARE, AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY (1996), also makes a critique of the balancing argument which is similar to that of this Article, but much more cursory. Boyle turns to theories of authorship as an interpretive construct. Again, assessment of the merits of this argument is beyond the scope of this Article.

I.B The Theory of Balancing

While the basic tension between providing incentives to create and restrictions on dissemination is intuitive, in order to specify the information needed to balance these factors, we must begin by describing the balancing approach in more detail. The balancing approach is an attempt to weigh the benefits of increased dissemination of works, against the restrictions on dissemination imposed by monopoly pricing:

> [C]opyright law seeks to establish a delicate equilibrium. On the one hand, it affords protection to authors as an incentive to create, and, on the other, it must appropriately limit the extent of that protection so as to avoid the effects of monopolistic stagnation.[23]

In principle the balancing approach could be applied in respect of individual works ("individual balancing"), for example by varying the term of protection in a manner which would provide each work with the minimum amount of protection needed to ensure its creation. This would obviously entail enormous administrative costs, and a more plausible version of individual balancing would simply grant or deny protection on a case-by-case basis.

Alternatively, and more prominently, the balancing approach can be applied to categories of copying or uses of a work ("categorical balancing"), so that certain uses of a work would be considered infringements, while other uses would be permitted. The most prominent example of this categorical balancing approach is in the defence of the idea/expression dichotomy, but it has also been raised in respect of translations and abridgments, which were at one point considered not to be infringements of the original author's copyright.[24]

This Part begins by considering categorical balancing, as it is more prominent, and then turn briefly to individual balancing. A balancing analysis can also be applied to categories of works, rather than categories of copying. For example, arguments that patent protection should be denied to algorithms because of the nature of the software development process and market tend to be of this nature.[25] I will not consider this form of categorical balancing simply because

---

[23] *Computer Assoc. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 696 (2d Cir. 1992).

[24] See Chafee, Reflections, *supra* n.000 at 511.

[25] See e.g. Pamela Samuelson, Benson Revisited: the Case Against Patent Protection for Algorithms and Other Computer Program-related Inventions , 39 Emory L.J. 1025 (1990). The argument against "software patents" also generally refers to the problem of an excess of invalid patents in the software area resulting from institutional

this article is concerned with arguments for the more traditional copyright doctrines, which apply to types of copying.

Categorical Balancing

Categorical balancing recognizes that a work of intellectual property is a basket of goods[26] which may be sold in different markets: a novel might be read by the final consumer, it might be dramatized or translated, or it might provide inspiration, ideas, facts, characters, plot lines, or a well turned phrase for use in other works. While denying copyright protection in one particular market will reduce incentives to create, it can also reduce the losses from restrictions on dissemination. If we choose the correct markets to deny and grant protection, net gains are possible on average. The central question is how to decide which markets should be protected. To do this, we need to know how denial of protection in any given market will affect both restrictions on dissemination and incentives to create.

The manner in which protection affects restrictions on dissemination is not always properly understood. For example, the Supreme Court in *Feist* criticised the "sweat of the brow" courts which had granted to protection to facts, on the basis that "they . . . declared that authors are absolutely precluded from relying upon the facts contained in prior works."[27] "In truth," the Court claimed, "'it is just such wasted effort that the proscription against the copyright of ideas and facts . . . [is] designed to prevent.'"[28] This argument is entirely misguided. We might as well say that protecting the market for derivative works absolutely precludes anyone but the author from writing a screenplay based on a novel. Protecting facts, or any other aspect of a work, does

---

limitations of the patent office: see e.g Samuelson *id.* at 1138. This aspect of the argument is closely to the property rights analysis which I develop in this article, since a plethora of invalid patents makes *ex ante* licencing more difficult. And whether it is couched in balancing or property rights terms, arguments about denying protection to classes of works, like software, are generally much more sensitive to the empirically difficult issues which are raises than are the traditional doctrinal balancing arguments.

[26] For convenience I will refer to the work as a whole as the "work" and the aspect of the work sold in individual markets as a "good".

[27] *Supra* n.000 at 1292.

[28] *Id.*. citing *Rosemont Enterprises, Inc. v. Random House, Inc.*, 366 F. 2d 303, 310 (CA2 1966), cert. denied, 385 U.S. 1009 (1967).

not preclude later authors from relying upon a prior work, it merely precludes them from relying on it without paying for the privilege.

In fact, restrictions on dissemination and increased incentives to create are two sides of the same coin. Contrary to the suggestion in *Feist*, the creator of a work has every incentive to ensure that the work is distributed as widely as possible, so long as licencing fees are paid. In the extreme, if the licencing costs, that is, the transaction costs associated with licencing, are zero, dissemination of the work will not be restricted, even though the licencing fees, that is, the price paid for the right to use the work, are positive and above marginal cost. If the creator of an original work could costlessly bargain with each subsequent user, the creator would never ask for such a high licencing fee that the second work would be unprofitable. The second author will not create the second work if it is not profitable to do so, and if the second author does not use the first work, she will pay no licencing fees. For the first author, some modest licencing revenues are better than no revenues at all. This is true even if the second work draws on the first substantially, but is marginal in the sense that it will not be profitable if more than purely nominal royalties were paid. Since the second author is left with positive profits, the second work will be undertaken. In other words, if licencing costs are zero, the first author will engage in perfect price discrimination, and, as is well known in the context of antitrust and regulated industries, pricing above marginal cost does not lead to a reduction in output if perfect price discrimination is possible.[29] If perfect price discrimination were possible in a given market, the question of whether to protect that market would be easy, since protection would lead to an increase in incentives to create without any restriction on dissemination. In that case protection would be clearly desirable.[30]

Of course perfect price discrimination is never possible, and the question of whether to

---

[29] Since the price to the marginal consumer equals marginal cost, the efficient output level is chosen: see Alfred E. Kahn, THE ECONOMICS OF REGULATION: PRINCIPLES AND INSTITUTIONS (1988, John Wiley & Sons) Vol.I at 131-133.

[30] The social benefit of a work includes the lowered cost of creation or value of earlier creation of a subsequent work; that is, its value as a building block: see Kenneth Arrow, Economic Welfare and the Allocation of Resources for Invention, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS, R.Nelson ed. (1962) at 617. If the author cannot capture this social benefit, the author's private incentive to create will be less than the social benefit, and some socially beneficial works will not be created.

protect is never so easy. The example of perfect price discrimination simply illustrates Coase's point that inefficiencies in the allocation of resources depend on transaction costs, not on the initial allocation of rights. In a more realistic setting where transaction costs are not zero, this implies that both the reward to the creator and the extent of restrictions on dissemination of the work depend on the particular pricing scheme used by the creator of the work.[31] The pricing scheme chosen by a profit maximizing author will depend on the transaction costs associated with various licencing schemes, which in turn depends on the institutions and technology of dissemination, as well as the shape of the demand curve for the good, and the uncertainty regarding the demand for the good.[32]

We also need to know the pricing scheme which will be used if protection is denied. We cannot assume that goods unprotected by copyright will be freely disseminated, since creators will seek alternative forms of protection, such as trade secret or contract. Since creators have an incentive to use the most efficient form of protection available and denying copyright protection reduces the available options, the alternatives will have equal or higher transaction costs than the copyright protection they replace. Net benefits from denying protection are therefore possible only if the alternatives are more limited in scope than copyright.

---

[31] This point is relevant to the significant literature addressing the question of the optimal tradeoff between patent length and breadth or scope given a fixed reward: see e.g. Pankaj Tandon, Optimal Patents with Compulsory Licencing 90 J.Pol.Econ. 470 (1982); Richard Gilbert & Carl Shapiro, Optimal Patent Length and Breadth, 21 RAND J.Econ. 106 (1990); Paul Klemperer, How Broad Should the Scope of Patent Protection Be? 21 RAND J. Econ 113 (1990). The basic argument is that "increasing the breadth of the patent typically is increasingly costly, in terms of deadweight loss, as the patentee's market power grows. When increasing the length of the patent, by contrast, there is a constant tradeoff between the additional reward to the patentee and the increament to deadweight loss..." Gilbert & Shapiro *id.* at 107. The assertion that increasing the breadth of the patent is typically increasing costly is true in the models used to calculate deadweight loss, but in practice it will depend on the actual pricing scheme used by the patent holder. Even imperfect price discrimination can greatly reduce the increment to deadweight loss as breadth of protection is increased, thus undermining the conclusion of the models. Another difficulty with this literature is the basic assumption that the scope of protection is assumed to be an adjustable policy instrument. In contrast I argue below that even if we knew what the optimal prize is, which we don't, adjusting the breadth of protection accordingly is beyond the competence of the courts. The scope of protection is determined by the cost of determining the bounds of the property right, rather than in order to fix the size of the reward.

[32] See Francis Bidault, TECHNOLOGY PRICING, (1989) at 28; F. J. Contractor, INTERNATIONAL TECHNOLOGY LICENSING, (1981) at 36, noting that uncertainty in demand and easier financing of payments are significant motivations for per unit royalties in practice. Under conditions of certainty regarding demand for the secondary good, the licence will be a fixed fee which extracts the entire surplus, and marginal cost will not be affected.

Nor can we assume that those goods which are unprotected under the alternative regime will be used by anyone who values them above marginal cost. There is a cost to distribution of the work, and if the work is unprotected it is more difficult to reap rewards from its distribution, so it is likely to be less widely distributed. Eliminating protection does not eliminate transaction costs, it just changes their nature.

Once we have determined the pricing and distribution systems both with and without protection, we need to derive the effect on incentives and dissemination of a change from one regime to the other. First consider the effect of changed incentives to create if protection is granted in a particular market. Knowing the demand curve and the pricing scheme which will be used tells us the appropriable surplus from this market, but to calculate the net benefit we need to know how this increase in revenue translates to an increased creation of works.

Most obviously, the increased revenue from that market increases incentives to create works generally. The number of additional works created for a given increase in revenue depends on the elasticity of supply for the work as a whole. The elasticity of supply for original works depends on where we are on the supply curve, that is, on total incentives to create, and total incentives to create in turn depend on which other markets are protected. So, for example, if only the market for original expression is protected, it might be that protecting the market either for translations or dramatizations would add significant incentives to create. But if the market for both expression and dramatizations were protected, that might be sufficient to draw in the great bulk of potential authors, and protecting translations in addition might have very little effect on the number of works created. If that were the case, protecting translations would be desirable if and only if dramatizations were not protected.

Further, the benefit from those additional works depends on how valuable those works are in each of the markets in which they can be sold. If we protect the market[33] for translations, for example, the benefit will be felt in an increase in original works, which flows into an increase in other kinds of derivative works, ideas and so on. In principle we need to know the net benefit in each of these markets in order to calculate the benefit from protection for translations.

---

[33] For convenience I will use "protect the market" to mean not just that copyright can subsist in the derivative work, but that the creation of a derivative work in that market infringes the copyright of the original creator

Granting protection will also affect the nature of the works created, since the author will have an incentive to change the composition of her works to favour goods which are protected at the expense of those which are not. This distortion in the composition of the work entails a deadweight loss since the private return to effort in creating the unprotected good is zero while the social return is positive. The goods comprising a work are produced jointly, which is to say that, for example, it is difficult to produce pure expression without producing any ideas. Because of this, output of the goods which are not protected will not be reduced to zero. The degree of reduction in the supply of a good which is not protected, and thus the welfare loss, therefore depends on the degree of jointness of supply between protected and unprotected goods.

To summarize, while denying protection to some categories of goods comprising a work can in principle increase net welfare, the balancing analysis needed to identify the goods to which protection should be denied requires considerable empirical information regarding the market structure: ideally we need to know the pricing strategy, transaction costs and shape of the demand curve at least for the market for the good in question, and preferably for all markets for the work; the degree of jointness of supply between the good in question and the work as a whole, and preferably the elasticity of supply of the work as a whole; the major alternative forms of protection and their associated transaction costs and pricing strategies.[34]

We should also note that the market factors are likely to be different in different sectors of the economy. There is no particular reason to think that the value of derivative works, or the methods of dissemination of works, is likely to be the same in the market for literature as that in the software market. This means that if the decision to protect the market for derivative works, for example, is to be defended on the balancing approach, we need to estimate the net effect of protection in all relevant markets, and then calculate a weighted average to decide whether protection is desirable.

Finally, dynamic effects must also be considered. One concern is that a policy based on a

[34] Note that the problem is formally the same when the question is whether protection should be denied to some class of works, such as software, rather than one of the goods comprising a work, except that non-intellectual property incentives such as the first mover advantage, rather than the jointness of supply between the goods comprising the work, provide the residual incentives which imply that output will not fall to zero if protection is denied.

static analysis will become outdated as markets evolve as a result of exogenous changes. For instance the advertising value of licencing fictional characters might increase as a result of global advertising, or the development of movies may have increased the demand for dramatizations of novels. Even more troubling, laws which were appropriate when introduced could impede the efficient evolution of markets. Transaction costs might be high in a thin market because low volume prevents the development of effective licencing schemes; or supply might be inelastic in an unprotected market for a certain type of work if there are only a few authors who are driven primarily by non-pecuniary factors. Present considerations might favour denying protection, and yet if protection were granted it might spur the development of innovative licencing schemes which would reduce transaction costs, or draw new authors into the market thus increasing elasticity of supply.[35]

I.C Categorical Balancing and the Idea/Expression Dichotomy

Direct empirical information on these issues is certainly not available to the courts in making decisions such as whether a work is sufficiently original to be protected, whether idea or expression has been taken, and whether protection should be denied to facts. But perhaps my analysis of the balancing problem has used economic jargon to make simple issues seem complicated. Maybe, the law of copyright makes "intelligent estimates" with sufficient confidence to defend the balancing approach.[36] We can only tell by considering the plausibility of specific balancing arguments. Accordingly, in this part I examine the balancing arguments commonly made in defending the idea/expression dichotomy in particular, and argue that they are inadequate.[37] The balancing argument does tell us that protection of all aspects of a work for an

---

[35] See Landes and Posner *supra* n.000 at 358; Robert P Merges, Of Property Rules, Coase, and Intellectual Property, 94 Colum. L. Rev. 2655 (1994).

[36] As Landes & Posner *supra* n.000 assert at 336.

[37] Breyer, The Uneasy Case for Copyright *supra* n.000 is a rare example of an attempt to apply the balancing argument in a detailed manner to determine whether protection of the market for expression of books is justified. He concludes that it may well not be justified. This illustrates the difficulty of the balancing argument in two ways. First, the fact that a detailed inquiry would lead such a prominent scholar to arrive at such a counter-intuitive conclusion in the core case of copyright in expression suggests that our initial intuition on issues such as ideas may also be suspect. Second, Breyer's article is ultimately unconvincing because, while he does undertake a detailed

unlimited duration would probably entail excessive licencing costs, but it does not tell us much more than this.

Doctrinally oriented justifications for the idea/expression dichotomy often apply the balancing metaphor rhetorically, but make no serious attempt to address the relevant issues. They simply point out that there exists a middle ground of protection which is preferable to the two extremes and then assert that the idea/expression dichotomy defines this middle ground. The Supreme Court's assertion that denying copyright protection to information and ideas "assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work,"[38] is a clear invocation of the balancing approach: ideas are denied protection to encourage dissemination, and expression is protected to provide incentives for creation. But this "building block" argument, as it has conveniently been termed,[39] is descriptively and normatively inadequate. Descriptively, a work can be used as a building block in a variety of ways, for example as a source of ideas, plot lines, facts, characters, snippets of expression, or as a basis for a derivative work. In some cases the subsequent use is an infringement, and in others it is not. And normatively, there is nothing inherently superior about using a work to create new works rather than for present consumption of the work: after all, the value of the future works lies in the fact that they are ultimately purchased as consumption goods. The simple fact that an aspect of a work is a building block for other works does not in itself imply that protection is or should be denied.

The building block defence of the idea/expression dichotomy cannot rest simply on the fact that ideas are used as building blocks. The claim must be something special about the use of ideas a building blocks. What might this be? Paul Goldstein claims that "[T]o give a creator a monopoly over these basic elements [ideas] would effectively stunt the efforts of other creators to

---

balancing analysis, he invokes no empirical evidence. In all, his article is a perfect illustration of my point that our intuitions are unreliable, and we do not have sufficient empirical evidence to arrive at firm conclusions.

[38] *Feist Publications Inc v Rural Telephone Service Co. Inc* 111 S. Ct. 1282 (1991) at 1290, citing *Harper & Row, Publishers, Inc. v. Nation Enterprises* 471 U.S. 539 105 at 556-7, S. Ct. 2218 at 2228-9. In a similar vein see Chafee, Reflections on the Law of Copyright, 45 Colum. L. Rev. 503 (1945) and NIMMER ON COPYRIGHT, §1.10[B].

[39] See Goldstein, *supra* n.000, §1.2.2.4.

elaborate on these elements in the production of their own works."[40] For Nimmer, the rationale for "exclusion of ideas from copyright protection is clear. To grant property status to a mere idea would permit withdrawing the idea from the stock of materials that would otherwise be open for development and exploitation."[41] It is true, as Nimmer states, that protecting ideas would "permit" withdrawing them from general exploitation, but he offers no reason to think that this would be the effect. As we have noted, authors have every incentive to licence the work widely, in order to maximize the revenues. Because of transaction costs, protecting ideas will reduce the use of those ideas by other creators, as Goldstein suggests, and perhaps this is what Nimmer had in mind. But why is this problem is any worse for ideas than it is for expression? Goldstein uses the words "effectively stunt" to imply the effect will be serious, but we are given no particular reason to think this is so, unless it is that ideas are "basic." But this label hardly mandates to any particular conclusion. If the basic elements of a work are more valuable to society, so that we lose more from restrictions on their dissemination, we must also lose more if they are never created in the first place.[42] Nimmer's reference to a "mere" idea is still more obviously an exercise in conclusory labelling. In arguments of Nimmer and Goldstein, denying protection to ideas seems clearly desirable only because they focus on restrictions to dissemination without mentioning the need for incentives to create.

If denial of protection to ideas is to be justified by a balancing analysis, we need to weigh both sides in that balance. And it may that ideas are indeed special in the sense that the gain

---

[40] *Supra* n.000 §2.3.1.1. Goldstein also emphasizes the argument that "To give creators a monopoly over such fundamental elements would reduce their incentive to elaborate these elements into finished works." id. This argument is quite misguided. It is true that if ideas are not protected, a creator will be forced to develop into protectable expression in order to realize any gain, but this is clearly undesirable if the creator of the idea is a great "idea person" with limited ability to transform the idea into expression. In that case, society would be better off if the work of elaborating on the idea could be contracted out to someone with more expressive ability but less ability at coming up with new ideas. And if the person who develops the idea is also good at expression, then she will have every incentive to turn the idea into expression into order to reap the additional rewards of the idea in the fully elaborated form.

[41] NIMMER ON COPYRIGHT, §13.03[B][2][a]. This argument is evidently derived from Melville B. Nimmer, The Law of Ideas, 27 S.Cal.L.Rev 119, 120 (1954).

[42] This is quite apart from the question of how to distinguish basic from non-basic elements of a work. Without any elaboration on this point, we may suspect that "basic" is a conclusory label rather than a reason for denying protection.

from reduced costs of new works is outweighed by the loss of incentives to develop new ideas.[43] But this is hardly a foregone conclusion, and assertion is an inadequate substitute for analysis. That ideas are building blocks for future works does not help answer this question, since the form of the loss, be it from reduced output of future works or reduced consumption by end-users, in itself says nothing as to the magnitude of the loss.

One functional difference between the use of ideas as building blocks as the use of other aspects of prior works is that the use of ideas is cumulative. When expression is borrowed, the source of the expression can be narrowly specified as a particular work. But in the development of ideas, each idea draws on a indefinite host of prior ideas. So, as Newton is reputed to have said, "If I have seen far, it is by standing on the shoulders of giants."[44] This is a powerful argument against perpetual copyright in ideas, since every Western thinker would then have to bargain with the heirs of thousands of thinkers since Plato for the right to use their ideas. But copyright is not perpetual.[45] When we take this obvious fact into account, the fact that each work builds on the thought of centuries supports rather than undermines the case for protection of ideas, since it implies that there are very few truly original ideas each generation. It is not obvious that the licencing of those few original ideas alone would be unduly onerous. This is simply a reflection of the general point made earlier, that the desirability of protection for one market depends on what other markets are protected. The balancing argument for denying protection to ideas becomes stronger as the duration of protection is extended, and it would be stronger yet if copying of non-original ideas was considered an infringement, but this merely

---

[43] M.B. Nimmer, Does Copyright Abridge the First Amendment Guarantee of Free Speech and Press? 17 UCLA L.Rev 1180 (1970) expressly recognizes at 1190-91 that "It is true that we have no positive evidence as to whether the flow would have been still greater had ideas *per se* been legally protectable, but there is reason to believe that idea protection would have in fact been counterproductive," but the only reason he offers is the building block argument itself, that is, the fact that "writers draw from the stock of ideas of their predecessors."

[44] M.B. Nimmer, Copyright and the First Amendment, *id.* at 1191, n.39 indicates that the phrase can be traced to R. Burton, THE ANATOMY OF MELANCHOLY (1618).

[45] Judge Easterbroook has remarked, *Nash v CBS Inc.* 899 F.2d 1537, 1540 (7th Cir. 1990) that "Every work uses scraps of thought from thousands of predecessors, far too many to compensate even if the legal system were frictionless, which it isn't." But given that the copyright term is limited, there is no reason to suppose that the number of predecessors who could claim compensation would number in the thousands. (Note, incidentally, that it follows immediately from the Coase theorem that Judge Easterbrook is wrong to assert that all predecessors couldn't be compensated if the system were frictionless.)

emphasizes the complex interrelationship of various market factors in the balancing argument.

Non-economic writing often describes the balance as being between the need to protect the author's incentive to create works and the First Amendment imperative of freedom of expression.[46] For example, Nimmer claims that "The market place of ideas would be utterly bereft, and the democratic dialogue largely stifled, if the only ideas which might be discussed were those original with the speakers."[47] But protection of ideas does not imply that the only ideas which might be discussed are those which are original to the speaker. In the first place, the vast majority of political ideas are ancient, and in the public domain. And for those few ideas which are original, pundits would normally be delighted to allow free use of their ideas, in return for the recognition and influence which that would bring. The notion that exchange of ideas would be "stifled" by protection is the reappearance of the fallacy that protection prevents distribution of works. There is certainly a real fear that political thinkers might try to licence their ideas narrowly, to prevent criticism, but the doctrine of fair use would clearly be applicable in such circumstances.[48] It is true that protecting ideas would be undesirable in a regime which granted perpetual copyright with no defence of fair use, but that it not our system. And more generally, this argument again makes denial of protection for ideas more attractive by focussing selectively on restrictions on dissemination rather than incentives to create. Ideas (or expression) which are not created because of inadequate incentives are as much a loss to the public welfare, however it is measured, as is restricted dissemination of speech which is created.

Those arguments which do address the relevant issues are not sufficiently detailed to be convincing. For instance, it is sometimes argued that protection for ideas is not necessary because ideas will be created even without complete protection, as the cost of coming up with

---

[46] See e.g. *Harper & Row Publishers v. Nation Enterprises* 471 U.S. 539, 560, 105 S. Ct. 2218, 2230 (1985); M.B. Nimmer, Does Copyright Abridge the First Amendment? *supra* n.000; R. Denicola, Copyright and Free Speech: Constitutional Limitations on the Protection of Expression, 67 Calif. L.Rev. 283 (1979) .

[47] NIMMER ON COPYRIGHT, §1.10[B][2].

[48] The economic case in favour of fair use for the purposes of criticism is well stated in the discussion by Landes & Posner *supra* n.000 of book reviews and parody at 358-359. On fair use and freedom of expression, see generally Denicola, Copyright and Free Speech, *supra* n.000.

new ideas is low, and authors will recover their costs from the market for expression alone.[49] This points to the fact that works are baskets of jointly produced goods, which is certainly relevant, and it is true that a large number of ideas are produced even without copyright protection. However, the simple fact of jointness of supply establishes little, as all the goods comprising a work are jointly produced to some extent. The important question is a more difficult one, namely the degree of jointness in supply between ideas and expression. That is, we need to know how many *additional* ideas would be produced as a result of protecting the market for ideas, as it is the value of these additional ideas which must be weighed against the restricted dissemination of ideas which will be produced without copyright protection. If anything, it seems likely that the degree of jointness of supply of ideas and expression is relatively low, since there is a plentiful supply of plot lines in the public domain which authors can draw upon, so that it is quite possible to write a successful novel with no new ideas.

Conversely, it is often suggested that protection is denied to ideas because there are few original ideas, or their value of most ideas is low, so that granting protection would provide little in the way of incentives to create additional works.[50] There are several problems with this argument. First, it may be that the reason that most works which rely primarily on copyright incentives for their creation are so lacking in ideas is that the authors cannot reap the benefit of any new idea which they might develop. That is, the low value of ideas in expressive works may be a result of, rather than a justification for, denying copyright protection.

Second, simply because the value of goods in given market is low, it does not follow that there is a net gain to be had from denying protection. If the value of the goods is low, the

---

[49] See Landes & Posner *supra* n.000 at 348 (argument (b)); Goldstein *supra* n.000 §2.3.1.1 "...copyright incentives are not needed for their [ideas] production."

[50] Goldstein, *id.* "This rule reflects the idea that [ideas as concepts] are relatively few and not worth the cost of monopoly protection." Some version of this also appears to be the primary argument made by Landes & Posner *supra* n.000 at 347-348 (argument(a)) as to why ideas are not protected. However, their argument is difficult to understand. They say that copyright protection for ideas would have a negligible effect on the cost of copying, and thus on the incentives of the author to create, because "Copiers are copying expression either unlawfully, in which case the marginal deterrence from protecting ideas, or lawfully, for example because their copying is deemed a fair use" *id.* at 348. But surely the main effect of protecting ideas would be on copiers who are now copying lawfully because they are copying ideas but not expression. The effect of protection on such copiers would not be negligible.

additional revenues from selling them is low, but the deadweight loss from users being priced out of the market is low as well. So, for example, Landes & Posner argue that the ideas used by a novelist are acquired "at zero cost, either from the observation of the world around him or from works long in the public domain."[51] But if this is true protecting those ideas imposes no additional cost on a subsequent author, who can acquire them costlessly on his own, without copying.

Finally, it is simply not true that the value of ideas is low. There is certainly a large body of work, academic writing, the value of which lies primarily in the ideas found therein, and not in the often turgid prose. There are at least some academic ideas the value of which far exceeds the value of any particular expression of that idea – the Coase theorem is a good example – and it seems clear that the returns from the sale of the expression alone are insufficient to provide an adequate incentive to create the ideas. That there are well-established mechanisms such as salaried academic positions, direct support for research, the Nobel prizes to encourage scholars to produce such ideas suggests that copyright protection for expression alone does not provide sufficient incentive to create such "idea works."[52]

To rebut this, it is suggested that protection should be denied to ideas because these alternative incentives to produce "idea works" are sufficient.[53] But this argument begs two important questions. First, how do we know that "enough" new ideas are being produced through these alternative incentives? To answer this question brings us directly back to the full

---

[51] *Supra* n.000 at 349-50.

[52] See Landes & Posner, *supra* n.3 at 350, acknowledging this point.

[53] Landes & Posner *supra* n.000 make this argument, albeit with a somewhat unusual twist. They argue, at 350, that "precisely because the [Coase] theorem is a powerful analytical construct, copyright protection would yield the inventor a very large income over and above the considerable nonpecuniary (as well as indirect pecuniary) income that accrues to a major theoretician. The total income would, in all likelihood, exceed the cost of inventing the theorem, thus creating a problem of rent seeking." This argument is extremely weak. The fact that Coase himself may have gained large rewards says nothing as to the expected return to theoretical work. Theory is much like writing novels – some strike it rich but most writers have day jobs. Most academics never in the careers come up with any truly new ideas, much less one as valuable as the Coase theorem. We might as well say that copyright should be denied to novels because John Grisham makes far more than the minimum necessary to induce him to write novels. Landes & Posner's reference to rent-seeking implies that they believe that if we protected original ideas there would be a mad rush into the field of legal theory in order to reap the rewards of inventing new ideas. This seems highly improbable. Note also that the problem of rent seeking does not arise when copyright is well defined: see *infra* n.000 and accompanying text.

complexity of the balancing analysis. And secondly, why should we suppose that these alternative incentives are preferable to copyright protection? The argument that public funding of public goods allows marginal cost pricing is an argument against intellectual property rights generally, and does not provide a basis for distinguishing ideas from expression. After all, government funding, such as the National Endowment for the Arts, also supports many works which *are* copyrightable, but we do not conclude from this that protection should be denied to expression. This argument also illustrates the problem that the balancing approach implies that the line between ideas and expression should shift as alternative incentives to create ideas change, for example as a result of changes in government funding for academic research, or conversely, as incentives to create works generally changes, for example when copyright protection for derivative works was established. Neither of these effects is apparent, although it is possible that they are present but subtle.

Economists have also argued that the cost of enforcing property rights in ideas would be high because they are ill-defined.[54] While the problem of poorly defined property rights is central to the approach which I develop below, the point that the cost of enforcing property rights in ideas is high is not sufficient when used to support a balancing argument since protection may be justified even though enforcement costs are high if the benefits of protection are correspondingly great.[55] Put another way, the enforcement cost argument is usually used to argue that the high cost of establishing property rights in ideas implies that property rights should be denied because the costs outweigh the benefits. In contrast, I am arguing that this cost-benefit analysis needed to

---

[54] Arrow, K. "Economic Welfare and the Allocation of Resources for Invention" in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY R.Nelson ed (1962); David Friedman, Standards as Intellectual Property: An Economic Approach, 19 U Dayton L. Rev. 1109 (1994); Landes & Posner *supra* n.000 at 349 (point (d)).

[55] We should also note an argument made by Landes & Posner which appears to be simply wrong. They argue, *supra* n.000 at 349 (point (c)), that ideas should not be protected because of the problem of rent-seeking: "Since the costs of developing a new idea are likely to be low in most cases relative to the potential reward from licencing the idea to others, there would be a mad rush to develop and copyright ideas. Resources would be sucked into developing ideas with minimal expression, and the ideas thus developed would be banked in the hope that a later author would pay for their use." This argument is unpersuasive in this simple form. While rent-seeking is a familiar problem in patent law, it is not a concern under copyright law if property rights are well defined: see discussion infra at n.000 and accompanying text. If a mad rush to develop new ideas occurs because the reward from licencing is greater than the cost of developing the idea, there is nothing wrong with this. It is simply the invisible hand of the market guiding resources to their most productive use.

come to this conclusion is itself not cost-effective because of the high cost of collecting the necessary information.

Of course all of these factors together may justify denying protection for ideas on a balancing analysis. But they may not. The point is that we do not have the information we need to decide. To suppose that a judge in any given case draws the line between ideas and expression in such a way that even roughly overall incentives would be provided if the same line were drawn in all other cases requires heroic assumptions regarding the ability of judges to draw accurate economic conclusions regarding overall trade-offs on the basis of intuition and evidence relating at most to a single work.

I.D Individual Balancing

If the need for information about the market structure makes categorical balancing infeasible, perhaps a balancing analysis which looked at individual works would have more manageable information requirements. Unfortunately, the informational requirements of the balancing approach cannot be avoided so easily.

For example, consider the approach advocated by John Wiley.[56] His basic argument is a straightforward application of the balancing approach: "Is the creation easy enough that considerations other than copyright protection suffice to motivate it, or is copyright's incentive a decisive condition of the creator's investment? If the former, then courts can declare open season on the work of authorship without imperilling the future supply of such new creations, thus giving consumers low prices from copiers."[57] But, somewhat unusually, Wiley's argument focuses on individual works: so, he argues that copyright law should ask whether ordinary experts in the field "believe that copyright law would have been superfluous to *them*, had they considered attempting *the work of authorship at issue.*"[58] While Wiley does not make the point expressly, this type of question appears only to require information about the particular work in

---

[56] John Shepard Wiley Jr., Copyright at the School of Patent, 58 U.Chi.L.Rev. 119 (1991).

[57] *Id.* at 148.

[58] *Id.* at 149, emphasis added.

question, which suggests it is plausibly within the competence of the courts.

But as with the balancing approach generally, the devil is in the execution. Consider how the test would be implemented. One way to implement the individual balancing text would be to deny protection to a given work if we determined, with the benefit of hindsight, that for that particular work non-copyright incentives would generate sufficient revenue to recoup the costs of creation. The problem with this approach is that the author does not have the benefit of hindsight in deciding whether to undertake a work. The incentive to create new works turns on the return which the author expects *ex ante*. Returns from successful works must be substantially greater than the cost of creation of those works in order to compensate for losses from unsuccessful works. To deny protection to successful works *ex post* simply reduces the author's expected return, and so has the same effect as an overall reduction in protection *ex ante*. This brings us directly back to the categorical balancing inquiry.

Alternatively, we might use a notional *ex ante* test, which appears to be what Wiley contemplated. But posing the question this way does not escape the problem. At the earliest stages of its inception, the expected return to any given work is simply the normal rate of return.[59] Applying Wiley's question at this point simply results in denying protection to all works or to none. At later stages the expected return to any given work will change. Some are more likely to succeed, while others will fail. If we deny protection to works which are likely to be successful, we have the same problem as applying the question with hindsight. But if we only deny protection to flops, the social gains are minimal.

Either way we approach it, from an *ex ante* perspective the individual balancing approach simply amounts to a general reduction in copyright protection and so entails the same informational requirements as categorical balancing.

I.E Evolutionary Arguments

An apparent counter-argument to the foregoing critique of the balancing approach is that a court need not have direct access to the relevant information in order to make efficient

[59] Some particularly talented authors will enjoy rents, but selectively denying protection to works by unusually successful authors will reduce the incentive for young authors who are not yet certain of their ability to enter the field. And of course such an approach is not remotely descriptively accurate.

decisions. It may be that the law evolves towards efficiency because of structural factors. Suppose judges have no information relevant to efficiency questions, and in general judges simply follow the rule laid down in previous case-law. Occasionally, and randomly from an efficiency perspective, the rule is reversed, say by a high court which is unhappy with the results when applied to the facts before them. But if inefficient rules, because they are out of line with the reasonable expectations of the parties, give rise to more disputes and get litigated more often than efficient rules, the law may plausibly evolve towards efficiency even in this informationally deprived decision making environment.[60]

I accept the general plausibility of the argument that the law can evolve towards efficiency. The property rights argument advanced in the next section is an argument that the law is efficient (although, as I have noted, it does not argue that the law is efficient in a strong sense).[61] The informational required for efficient decision making is nonetheless relevant.[62] In particular, the evolutionary argument from litigation pressure does not plausibly explain efficiency in areas involving the application of a standard, such as the idea/expression dichotomy, fair use, and the boundary between copyright and patent. Some digression is required to explain why.

While economists are commonly attacked for unrealistically assuming perfect rationality, the whole enterprise of economic modelling is premised on bounded rationality. As Mitlon Friedman argued in a well-known article, economic models do not precisely reflect social

---

[60] The classic statement of the evolutionary defence of the rationality assumption in economics is Armen Alchian, Uncertainty, Evolution and Economic Theory" 58 J. Pol. Econ. 211 (1950). Much of the impetus for the debate in law and economics is the claim of Richard Posner that the common law is efficient: see generally Richard A. Posner, Economic Analysis of Law, 4th ed. (1992). The earliest model evolution of the common law is Paul Rubin, Why is the Common Law Efficient, 6 J. Legal Stud. 51 (1977). The debate as to whether structural pressures will lead to efficiency in judge made law remains unresolved: see the review by Paul Rubin, Judge Made Law, Section 9200, Encyclopaedia of Law and Economics (forthcoming).

[61] I do not draw a distinction between common law and statutory law in this respect. Many of the arguments which suggests that the common law may be efficient may also be applied to statutory law: see Donald Wittman, Why Democracies Produce Efficient Results, 97 J. Pol. Econ. 1395 (1989).

[62] Information considerations in a broad sense are well recognized as being relevant to the evolutionary argument. For example one objection to the structural argument for efficiency of judge made law is that the information generated by structural pressures is biased: see Gillian Hadfield, Bias in the Evolution of Legal Rules, 80 Georgetown L.J. 583 (1992). My argument is that the information directly available to the courts is also relevant.

reality.[63] A model which did fully reflect the complexity of social reality would be too unwieldy to yield useful results. This argument applies to decision making generally, and to judicial decision making in particular. Human rationality is limited, so we must make our decisions on the basis of a model which simplifies reality. The model does sacrifice something in accuracy, but if the model is a good one, what is lost in detail is more than compensated for by increased tractability. A decision making model is a good one if it is simple enough, in terms of the criteria which it deems relevant to the decision, and the relationship between them, to be tractable, given our limited human rationality, and yet these factors are sufficiently good proxies for the underlying complex reality that the model provides good results.

In traditional modelling we construct the model so that (we hope) it captures some important aspect of reality. Because we construct the model, we understand why the particular rules applied to the criteria in question determines the efficient outcome. The evolutionary argument suggests that workable models can arise even if the decision maker does not understand why the model works. This allows for the construction of more efficient models for any given degree of rationality on the part of the model builder. But the model must still be applied to the facts by a decision maker with limited rationality. This means that even if the model is evolved rather than consciously constructed, the trade-off between accuracy and tractability still arises in the application of the model. If the model is to provide good results, it must either incorporate factors which are directly relevant to a determination of the efficient outcome on the facts, or it must incorporate good proxies for those factors. This is true whether or not the decision maker understands the model.

The informational argument against the balancing approach to the idea/expression dichotomy can be restated as saying that information regarding the factors directly relevant to an efficient decision on the facts of any particular case is not available, and there are no good proxies for these criteria. Even granted that a judge need not understand why the application of particular criteria lead to efficient decisions, an efficient decision is not possible if the relevant information is not available. The point is highlighted by the fact that the balancing approach to

---

[63] M. Friedman, The Methodology of Positive Economics, in M. Friedman, ESSAYS IN POSITIVE ECONOMICS (1953).

the idea/expression dichotomy implicitly defines "idea" as that which it would be inefficient to protect. This is clearly not a good proxy for underlying factors relating to efficiency.[64]

In contrast to "ideas", "facts" and "translations" for example, are sufficiently well defined outside of the balancing model that they may serve as proxies for factors related to efficiency. It is therefore plausible that rules regarding protection of facts or derivative works could evolve towards efficiency, provided that the evolutionary pressures do in fact tend to encourage efficiency. But even in this context, informational requirements are important. A decision making goal which requires a great deal of information will require a more complex model to achieve a given degree of accuracy than one which requires a small amount of information. Conversely, for any given degree of tractability, a model for achieving an informationally demanding goal will be less accurate than one which seeks to achieve a less difficult objective. Again, this is true regardless of whether the model evolves or is consciously constructed. If we have alternative possible goals, it may therefore be preferable to seek to achieve the simpler goal with some accuracy, rather than achieving the more difficult goal poorly. Certainly when a court is required to consciously apply a model, for example when deciding a controversial issue, it should apply the simpler model.[65]

PART II THE PROPERTY RIGHTS APPROACH

The balancing approach aims for an optimal initial allocation of property rights. This is desirable, but very ambitious. The property rights approach suggests that we set our sights on a more modest goal, namely to ensure that intellectual property rights are clearly defined. This is done by denying protection when granting protection would result in poorly defined IP rights. More specifically, protection should be denied when granting protection would result in a significant chance of erroneous findings of infringement. Since this reduces incentives to create

---

[64] See Mario J. Rizzo, Law and Flux: The Economics of Strict Liability in Tort, 9 J. Legal Stud. 291 (1980), esp. at 317, making a similar argument in favour of strict liability rather than negligence; and see Peter H. Aranson, Economic Efficiency and the Common Law: A Critical Survey, in von der Schulenburg & Skogh, eds. Law and Economics and the Economics of Legal Regulation (1986) esp. §3.2 for a survey of the literature which emphasizes this point.

[65] See the discussion of the Supreme Court's decision in *Feist infra* Part III.A.

the works to which protection is denied, costs and benefits must be weighed. While this comparison is not trivial, it is much easier than that which is required by the balancing approach. I will argue that, in contrast to the balancing approach, reasonable guesses can be made regarding the property rights trade-off on the basis of the basis of information available to the courts.

Poorly defined intellectual property rights may take the form of underprotection of prior works, that is, judicial findings that there was no infringement when in fact there was; or overprotection of prior works, that is, erroneous findings of infringement when there was none.

Underprotection of the work is not a serious problem. In the context of tangible property underprotection of individual rights is likely to lead to overexploitation of the resource (the tragedy of the commons). In contrast, since intellectual property is subject to non-rivalrous consumption, overexploitation is not possible. The only consequence of underprotection is that incentives to create are smaller than if infringement findings were error free. While gross underprotection of intellectual property will seriously weaken incentives to create, marginal errors in the direction of underprotection will have only a small effect. Certainly limiting the scope of protection, for example by denying protection to ideas, cannot be justified as a means of preventing false negatives, since incentives to create will be smaller yet if protection is denied.

The possibility of erroneous findings of infringement, on the other hand, can inhibit the production of new works and result in losses which are much greater than the value of the work which is being protected. Because intellectual property rights are property rights, a successful plaintiff armed with the threat of an injunction can potentially extract the defendant's gross revenue *not* net of sunk costs.[66] To protect herself against this, the creator of a new work must licence prior works which she might be found to infringe before incurring significant sunk costs in creating her own work. In general, this means that property rights encourage *ex ante* bargaining over the rights, which allows for better allocation of rights than *ex post* allocation by

[66] J. R. Green & S. Scotchmer, On the Division of Profit in Sequential Innovation, 26 Rand J. Econ. 20 (1995), and S. Scotchmer, Standing on the Shoulders of Giants, 5 J. Econ. Perspectives 29 (1991) emphasize this problem and discuss joint ventures as a solution in the context of patent law.

the courts, as with a liability right.[67]

But *ex ante* bargaining is only possible if property rights are well defined in the sense that the creator of the work is able to determine *before* she acts whether she will be found to have infringed a prior creator's rights. So, for example, copyright is well defined if the author of a new work knows that the court will not find infringement unless she in fact copied from a particular earlier work. Since copying is normally conscious, the creator can then protect herself by licencing from works which she knowingly used.[68] In contrast, when erroneous findings of infringement are likely an author undertaking a new work must be concerned about the possibility of subsequent actions by creators whose works she did not in fact copy from. This class of prior potential plaintiffs is large and indeterminate, and as a result the author cannot confidently protect herself by *ex ante* bargaining. I will refer to actions by plaintiffs whose works were not in fact copied as "opportunistic" lawsuits, since the problem arises because of the presence of sunk costs and the difficulty of complete *ex ante* licencing.[69]

The author contemplating a new work can take account of the threat of opportunism by licencing from those prior authors who she did not copy from, but who might successfully bring an opportunistic suit, or by discounting her expected returns by the possibility of a successful opportunistic action, or most probably, by doing both. The increased scope of licencing will increase transaction costs and may prevent some works from being undertaken. And more importantly, because a successful opportunistic action can extract net profits not net of sunk

---

[67] The classic article is Calabresi and Melamed, Property Rules, Liability Rules and Inalienability: One View of the Cathedral, 85 Harv.L.R. 1089 (1972). Louis Kaplow and Steven Shavell, Property Rules Versus Liability Rules: An Economic Analysis, 109 Harv. L. Rev. 713 (1996) try to turn this on its head, arguing for example that "even when courts are uncertain about the magnitude of harm, liability rules are superior to property rules" (at 719). While a full critique of the article is obviously out of place here, I find their thesis unpersuasive except in some special circumstances.

[68] Unconscious copying of expression constitutes infringement, notwithstanding that the defendant believed in good faith that the work was the product of his own imagination: see e.g. *ABKCO Music Inc. v Harrisongs Music, Ltd.*, 722 F.2d 988, 998 (2nd Cir. 1983). The obvious justification for this doctrine is to facilitate proof of infringement. It does in principle create some undesirable uncertainty over prior licencing, but in practice it is probably sufficiently rare that it can be ignored, and the benefits gained in terms of reduced litigation costs justify the doctrine.

[69] This corresponds with the use of the term by Oliver Williamson, THE ECONOMIC INSTITUTIONS OF CAPITALISM (1985).

costs, the discount to allow for the threat of opportunism may make expected returns negative, in which case the creator will choose not to undertake the work. (These issues are discussed in more detail below.)

These problems are ultimately manifestations of waste from rent-seeking. The rent-seeking arises because the threat of opportunism means the subsequent author must allow for payments to a prior author whose work she did not actually use, which means that the private value of the prior author's work is greater than it's social value. The rent-seeking problem which arises as a result of poorly defined copyright is very similar to the well known patent race problem. The following discussion therefore begins by examining the problem in the context of copyright, to set the stage for a comparison of patent and copyright. The main source of poorly defined copyright arises when independent creation of a work is likely, but is difficult to establish. In that case similarity does not serve as *prima facie* proof of copying, and false findings of infringement would be likely because establishing copying by direct evidence is unreliable. Both the originality requirement and the idea/expression dichotomy can be explained as denying copyright in these circumstances. I will focus first on the originality requirement, as this allows an elaboration of the basic argument while postponing the issue of the relationship between copyright and patent. That question is then taken up in the subsequent section on the idea/expression dichotomy. In that section I will argue that when copyright and patent are both well defined, copyright is superior to patent, in which case copyright protection should be used. When patent is well defined and copyright is poorly defined, patent protection is desirable. When neither is well defined, protection should be denied entirely.

II.A The Originality Requirement in Copyright Law

The requirement that a work be original in order to attract copyright protection means that exact reproductions of public domain works will not be protected, while less faithful reproductions will be, even though the variations which attract protection result from poor craftsmanship or technical limitations in the reproduction process rather than artistic choice.[70] This presents a puzzle, since the doctrine applies even in cases in which the value of the

[70] See Wiley, *supra* n.000 for a good discussion of the cases.

reproduction to consumers lies in fidelity to the original, so that the variations needed to attract copyright protection are undesirable. Wiley captures the problem in referring to "the perversity of defining originality as variation."[71]

A balancing approach to the originality requirement, such as that proposed by Wiley, suggests that exact reproductions are not protected because their value is minimal, and denying protection reduces deadweight losses associated with licencing reproductions which are created even without copyright protection.[72] This is not persuasive, either descriptively or normatively. In the first place, if the value of reproductions is in fact low, then the benefit of increasing access to them by denying protection is minimal. Moreover, it simply is not true that the value of reproductions is generally low. An exact reproduction of a public domain work may well be valuable in making the work more accessible. Further, the market is the best judge of value: if reproductions were protected a second author would not buy a licence unless the cost of doing so were less than the cost in time and inconvenience of going to the public domain work. The fact of copying is in itself evidence that the reproduction was valuable. As one court put it, "after all, there remains the rough practical test that what is worth copying is worth protecting."[73] It is precisely because the courts wish to leave the assessment of value to the market that the threshold for originality is low and does not require a showing of artistic merit.[74] And descriptively, the law is clear that original works which are not valuable may be protected, while reproductions which may be very valuable will be denied protection. This is inconsistent with the balancing approach.

The real benefit to denying protection to exact reproductions is not increased dissemination of the reproductions, but rather increased dissemination of the *original*. As Judge Posner noted in *Gracen v Bradford Exchange*, because of the difficulty a subsequent creator

---

[71] *Id.* at 133.

[72] *Id.* at 148.

[73] *University of London Press Ltd. v University of Toronto Press Ltd.* [1916] 2 Ch. 601, 610 per Petersen J.

[74] See *Bleistein v Donaldson Lithographing Co* 188 U.S. 239, 251 (1903). For a recent discussion and affirmation of the low standard for originality, see *CCC Information Services, Inc., v. Maclean Hunter Market Reports, Inc.,* 44 F.3d 61 (2nd Cir., 1994) esp. at 66.

would have in showing that he copied from the original rather than a reproduction, protection for reproductions "would paradoxically inhibit rather than promote the creation of [derivative works] by giving the first creator a considerable power to interfere with the creation of subsequent derivative works from the same underlying work."[75] This spare comment captures the essence of the property rights argument. Considerable elaboration is in order, since Judge Posner's argument has been challenged on its merits, and because I will argue that parallel reasoning explains the denial of copyright to ideas.[76]

Copying is normally established by showing substantial similarity and proof of access, but substantial similarity to a reproduction may simply indicate independent copying from the original. This is so even if the defendant had access to the plaintiff's work, since the defendant may nonetheless have chosen to copy the original. Putting the onus on the plaintiff to show by direct evidence that the defendant did copy from its reproduction is tantamount to denying protection altogether. On the other hand, if the onus is on the defendant to prove by direct evidence that it did not copy, occasional false findings of infringement are inevitable. Protection for exact reproductions would therefore mean that a prior reproducer would have a significant chance of success in an infringement action against a subsequent creator who had copied not from the plaintiff's reproduction, but from the original or from a different reproduction. To protect herself from the possibility of erroneous findings of infringement the author of a new work would have to search for and licence all prior reproductions of the original to which she had a reasonable opportunity of access. This will not always be feasible, and so the author wishing to use an original would generally have to factor into her calculations some provision for expected cost of a successful suit. This cost could make use of the original infeasible.

Wiley rejected the *Gracen* court's argument regarding difficulty of proof, arguing that if this were the real problem, "the appropriate solution is a high standard of proof for plaintiffs."[77]

---

[75] 698 F2d 300, 305 (7th Cir. 1983).

[76] The concern raised here is quite distinct from that raised by Landes & Posner, *supra* n.000 at 000 who noted that the cost of new works will increase because of the need to pay licence fees to creators of previous works which are used in the new work. In contrast the issue raised by the property rights analysis is that the creation of new works which do *not* incorporate the previous work may be impeded.

[77] *Supra* n.000 at 137.

But if the appropriate standard of proof is so high that the plaintiff can never win, this is tantamount to holding that reproductions are not protected, in which case we might as well, for the sake of clarity, deny protection as a matter of law. In the subsequent discussion I will argue that the approriate standard is indeed so high that the plaintiff should never win. Wiley also argued that the *Gracen* court "embraced the argument in a case in which it did not apply; the court expressly identified the facts that made plausible its summary judgment supposition that the [defendant's work] was indeed a 'piratical copy' of [the plaintiff's work]."[78] But it is not enough that it is simply plausible that the defendant copied the plaintiff's work; rather the question is whether copying can be established with sufficient confidence to preclude the threat of opportunism, which is a much higher standard.

Du Puy v. Post Telegram Co.[79] illustrates the problem. The plaintiff was a journalist who claimed copyright in his article on the history and significance of Peace Day which he had sent to a number of newspapers before it was finally published in the *Washington Star*. He had copied the article almost verbatim from a bulletin which had been widely circulated by the Bureau of Education. The defendant newspaper had also published an article on Peace Day which also contained much of the exact wording of the bulletin. It seems likely that the defendant copied from the original bulletin, but, so far as we can tell from the facts, it is possible that it copied from the plaintiff's article. If reproductions were protected, then the court would have had to attempt to determine the source of the defendant's article on the facts. The usual best evidence of copying, substantial similarity, is of no help. The plaintiff journalist can be expected to bring suit so long as the expected return is greater than his costs. With relatively large rewards to a successful suit, the probability of success need not be particularly high. At the same time, it is clearly impossible for the defendant newspaper to protect itself by licencing from every journalist who wrote a story based on the Peace Day bulletin. Even if the probability of a successful action by any individual plaintiff was low, there are no doubt many journalists who wrote, or would be tempted to write, similar articles, and the likelihood of a successful action in at least one case would be appreciable. In the worst case, unless the plaintiff is denied copyright entirely in his

[78] *Id.*.

[79] 210 F. 883 (3rd. Cir, 1914).

reproduction of the bulletin, newspapers will choose not to publish anything at all which is based on the Peace Day bulletin.

The worst case loss from granting protection to reproductions is clearly greater than the worst case loss from denying protection. In the worst case, if protection is granted the full value of the original in all subsequent works is lost because of the fear of opportunistic infringement actions by reproducers. In contrast, the worst case loss from denying protection arises when the incentive to create reproductions derives entirely from copyright protection, in which case no reproductions will be created if protection is denied. But the benefit of reproductions lies only in the reduced cost of access to the original, and this obviously cannot be greater than the value of the originals themselves. Whether protection is granted or denied to reproductions, the original can still be used for consumption purposes, at the full access cost, but when protection to reproductions is denied, the original can also be used in subsequent works other than reproductions. This continued ability to use the original in subsequent works is the net benefit of denying protection to reproductions.

In less extreme situations, where the opportunistic value of the reproductions is less than the full value of the original in subsequent works, the question is more difficult. Protection for the reproduction is unlikely to entirely preclude use of the original in all cases, and denying protection to reproductions has its own costs, in the form of an under-supply of reproductions. If the threat of a false finding of infringement if reproductions are protected is sufficiently low, then the benefit of having the reproduction may outweigh the harm of under-use of the original.

This raises a balancing question under the property rights approach, but the nature of the balancing task is quite different from that which arises under the balancing approach. To see why, consider first the problem of restricted use of the original in more detail. If property rights are poorly defined but reproductions are copyrightable, an author who wants to use an original (either directly or by licencing the right to copy a reproduction) faces a threat of opportunism from prior reproducers who she did not in fact copy from. To protect herself from this threat, the author will licence from prior reproducers, to the extent that it is cost-effective to do so, and then discount her expected return from the creation of the work to allow for the expected cost of an opportunistic action. The cost of the reproductions to subsequent authors, which, conversely, is

the benefit to reproducers, has three components: the fee paid for use of the reproduction for the purpose of reducing the cost of access to the original; the fee before the creation of the subsequent work to reproducers whose works were not used, in order to forestall opportunistic actions (the "*ex ante* opportunistic cost/value"); and the expected cost of opportunistic actions *ex post* (the "*ex post* opportunistic cost/value").

Of course neither *ex ante* nor *ex post* opportunism value represents a social benefit, and because the private value exceeds the social value, a rent-seeking problem arises. However, the distortions engendered by rent-seeking manifest themselves differently in the two cases.

If reproducers can price discriminate perfectly, *ex ante* opportunism will not directly decrease the number of subsequent works which are created.[80] It will increase the returns to the creation of reproductions, and so a wasteful excess of reproductions will be created, as one would expect in a rent-seeking situation. And because reproducers cannot price discriminate perfectly, the increased scope of licencing necessary to forestall opportunistic actions will increase transaction costs and reduce the number of subsequent works which are created. There will be some offsetting effect as the fee paid for reproductions which actually are used declines, because there are more reproductions. The precise net effect depends on the same market factors discussed in the context of the balancing approach.

In *ex post* opportunism, in contrast, the waste is not manifested in the creation of excess reproductions, as is typical when rent-seeking is a problem. *Ex post* opportunism represents a minimum cost of using the original or a reproduction, which cannot be negotiated away *ex ante* by a single reproducer, even if perfect price discrimination by that particular reproducer were possible. A subsequent author who is deciding whether to use the original or a reproduction will take into account the fact that this will expose her to *ex post* opportunism, and so will factor in this opportunism cost as part of the cost of using the original. On average, for those authors who choose to use the reproduction, this will reduce the amount which she is willing to pay any reproducer for the right to use the reproduction by exactly the same amount as reproducers generally can subsequently extract through opportunism. There is no net benefit to reproducers, so the number of reproductions will not increase.

---

[80] See discussion *supra* n.000 and accompanying text.

But *ex post* opportunism is not simply a wealth transfer with no efficiency implications. A net loss occurs because for some authors the opportunism cost will be higher than the value of using the original or a reproduction, and the author will choose not to use to the original at all. The subsequent author is harmed, and there is no offsetting gain to the reproducer, resulting in a net social loss. As the cost of *ex post* opportunism increases, fewer authors will choose to use the original (again, either directly or by copying a reproduction). This tends to reduce the number of reproductions, both by reducing the value of the subsequent works which might either use the reproductions or be the subject of opportunism, and because reproductions are themselves subsequent works which are subject to opportunism. In the worst case, no subsequent works of any kind which use the original, including reproductions, are created, and the maximum loss, which is the loss of the value of the original in all subsequent works, is realized. Thus *ex post* opportunism has a more direct effect on the number of subsequent work created than does *ex ante* opportunism because the effect of *ex ante* opportunism is diluted by the possibility of *ex ante* bargaining. The effect of *ex post* opportunism does not depend on the structure of the market for the works in question.

Having specified the nature of the harm imposed by opportunism, the next question is its extent. The maximum opportunistic value of reproductions depends on the value of subsequent works and the probability of an erroneous finding of infringement. The balance between *ex ante* and *ex post* opportunism depends on how wide a scope of *ex ante* licencing is possible. The balance will tend to tilt in favor of *ex post* opportunism because of an adverse selection problem. The author of a subsequent work makes the *ex ante* licencing decisions on the basis of expected revenues, whereas the reproducer's decision to sue *ex post* is made on the basis of realized revenues from the defendant's work. The more successful the second work, the lower the probability of success which will induce a prior creator to sue. Further, *ex post* search costs by the prior author are likely to be lower than *ex ante* search costs for the creator because the prior author is more likely to bring an action when the defendant's work is successful, so search costs will be low, while the prior author's own work may be relatively obscure. So, the user may choose not to licence *ex ante* because the expected revenue from the work does not warrant it, but the discount to allow for *ex post* opportunism may still be large because an *ex post* suit is

most likely in precisely those cases where the work is unexpectedly successful. This problem is always present to some extent as there is always some uncertainty in the expected success of a work, but it will be especially severe if the economics of the industry are such that a small number of successful works subsidizes a large number of flops. In that case a relatively small probability of false finding of infringement may have a large effect on incentives to create.

Once we have determined the *ex post* opportunism cost, we need to relate it to the reduction in the number of subsequent works. This depends on the distribution of the producer surplus generated by use of the work. If the work adds a small amount to many works, for a significant net surplus, even a small *ex post* opportunism cost will drastically reduce the use of the work. In contrast, if few subsequent works use the work in question, but each has a relatively large surplus from the use of the work, a small *ex post* opportunism cost will have a negligible effect on welfare.

So, in principle, the cost of denying protection to reproductions depends on several factors: the probability of a false finding of infringement, which determines the opportunistic value of the reproduction a proportion of the value of the original in subsequent works; the uncertainty in the demand for the subsequent work, which determines the balance between *ex post* and *ex ante* opportunism; and the distribution of the producer surplus generated by the use of the work, which determines the reduction in the number of works as a result of *ex post* opportunism.

The first of these factors is clearly within the competence of the courts to decide. The only factor expressly taken into account by the courts, whether there is any variation in the aspect of the work for which protection is claimed, is an excellent proxy for the probability of an erroneous finding of infringement. On the other hand, the latter two factors, relating to the precise magnitude of the loss from opportunism, cannot readily be determined on the basis of the available evidence, and the cases give us no reason to believe that the courts attempt to account for them. So, it would appear that of the relevant factors, the courts adopt a strategy of necessity in taking into account only the probability of a false finding of infringement.

Is this strategy of necessity a reasonable one? Can the courts safely focus on the probability of a false finding of infringement, to the exclusion of other factors which are relevant

in principle? In most cases the answer is yes. To see why consider the other side of the equation, namely the benefit of granting protection to reproductions. The value of allowing reproductions lies in the reduced cost of access to the original. The more accessible the original, the lower the value of allowing access to it. When the original is readily accessible, the issue is easy to decide, even in the absence of information regarding the precise effect of opportunism. We can simply assume the worst, that denying protection to reproductions results in no reproductions being produced. Even in that worst case, all that is lost is the increased access to the original provided by the reproductions, and when the original is widely available already, this loss is minimal. On the other hand, the worst case loss from granting protection is the value of the original in subsequent works, which is much greater. The court, therefore, should not grant protection unless the reproduction is sufficiently different from the original that the court is confident that copying from the reproduction can be distinguished from copying from the original by a court in an infringement action.[81] Of course, this inquiry is not mechanical. Without denying protection entirely to all works, it is not possible to ensure that the probability of overprotection is zero. And we can't quantify how low the probability of overprotection has to be before it is low enough. It just has to be low enough that subsequent users of the original don't have to worry about it. This is not a precise number, but it is nonetheless provides a reasonably certain guideline.

The problem is more difficult in the context of what might be termed "archeological" reproductions of a rare original which make the work much more readily available. In the extreme, the cost of access to the original might be so high that no one would use it in creating new works, which means that the worst case loss from denying protection to reproductions could be just as high as the worst case loss from the opportunism caused by granting protection. But the evidentiary focus of the property rights analysis suggests that the threat of opportunism is likely less in the case of archeological originals. A defendant who in fact copied from the

[81] Note that as in the balancing approach, the fact that reproductions are often low in value is relevant, but in the property rights analysis denying protection to low value works creates a net benefit by encouraging access to more valuable originals. In contrast, in the balancing approach denial of protection to low value works is said to be beneficial because of increased access to the low value work itself. This is a different argument, which is subject to the critiques made *supra* n.000 and accompanying text.

original should be able to prove it relatively easily, because the trouble taken to gain access to the rare original will have left a trail. The defendant who knows the rule can ensure this by taking steps to establish evidence of access to the original. On the other hand, a rule which granted protection for archeological reproductions but not for reproductions of easily accessible originals raises a potentially difficult line problem of distinguishing between the two types of originals. The real choice therefore is between the lower administrative costs of a clear rule denying protection to all reproductions and the benefits of a more costly rule which offers protection to some archeological reproductions. The choice between the two rules is not clear, and not surprisingly, the few decided cases are not entirely consistent. Most prominently in *Alva Studios, Inc. v. Winninger,*[82] the Court held that an exact scale reduction of Rodin's famous "Hand of God" was protected, and this decision was later approved by the Second Circuit (en banc) in part on the basis that "Rodin's sculpture is. . . so unique and rare, and adequate public access to it such a problem that a significant public benefit accrues from its precise, artistic reproduction."[83] We might also suggest that access to the original is so tightly enough controlled that the court could be quite confident on the facts that the defendant had not copied from the original.

[82] 177 F. Supp. 265 (S.D.N.Y. 1959).

[83] *Batlin & Son v Synder* 536 F2d 486, 492 (2nd Cir) (en banc) *cert denied*, 429 U.S. 857 (1976). The court also cited the skill of the reproducer (at 491), which also implicitly acknowledges the value of the reproduction in making the original available. The other main reason for approving the finding that the reproduction was original was its "exactitude" (at 491), and the court indicated that if the model in issue in *Batlin* might have been protected has it been an "exactly faithful reproduction" (at 492). This is quite astonishing, as it appears to directly contradict the *Batlin* Court's direct statement at 491 that "that to support a copyright there must be at least some substantial variation, not merely a trivial variation such as might occur in the translation to a different medium." Compare *Hearn v. Meyer* 664 F. Supp. 832, 1987 U.S. Dist. LEXIS 6552 (1987, SD NY) in which the plaintiff had tracked down and labouriously reproduced rare illustrations from the original edition of the Wizard of Oz and the court nonetheless held that the reproductions were not sufficiently original to attract copyright.

II.B The Idea/Expression Dichotomy

The idea/expression dichotomy is closely related to the originality requirement. Most obviously, one common explanation of the idea/expression dichotomy is that the ideas are denied protection because of lack of originality.[84] This is often persuasive. While the courts may insist that an idea will be denied protection regardless of originality, in many cases the ideas to which protection is denied are, on the facts, almost certainly not original. In these cases denying protection on the basis of the idea/expression dichotomy rather than on the basis of the originality requirement can be explained simply as a method of avoiding the costs of a formal inquiry as to originality when the outcome is a foregone conclusion.[85]

But this cannot be the whole story, as cases do exist in which protection is denied to an idea when originality is a real issue. Original but uncopyrightable ideas may be divided into two categories: works which are more appropriately protected by patent ("potentially patentable ideas")[86]; and abstract ideas, which are protected by neither patent nor copyright.[87] Denial of copyright to both these classes of ideas can be justified on essentially the same basis as the originality requirement. Copyright in ideas, like copyright in reproductions, would be poorly defined because if ideas were copyrightable, the defence of independent creation of the idea would be plausible, but difficult to establish.

---

[84] E.g. Wiley *supra* n.000; Alan T. Dworkin, Originality in the Law of Copyright, 39 Boston U. L.Rev. 526 (1959); Goldstein *supra* n.000, §2.3.

[85] *Nichols v Universal Pictures Corp.* 45 F.2d 119 (2nd Cir. 1930) is an example. Hand J. indicated at many turns that what had been taken was not original to the plaintiff, while insisting that the result would have been the same even though it was. For example, Hand J. states at 122 that "We assume that the plaintiff's play is altogether original, even to an extent that in fact is hard to believe," and he makes similar remarks on the same page regarding the originality of the plot and characters.

[86] A note on terminology. Goldstein *supra* n.000 §2.3.1.1 refers to what I have called potentially patentable ideas as "ideas as solutions". Note that while it is "hornbook law" that "an idea of itself is not patentable" (*Gottschalk v. Benson*, 409 U.S. 63, 67 (1972)), it is also true that some aspects of a work which may be unprotectable by copyright on the ground that protection is being sought for an "idea" may nonetheless be protectable by patent. The rule that abstract ideas in the patent sense cannot be patented is discussed *infra* n.000 and accompanying text. For the moment it is enough to say that the term "idea" is being used in two different senses and I am using "ideas" in the copyright sense. The assertion that "ideas" (in the copyright sense) are more properly protectable by patent is not inconsistent with the rule that "ideas" (in the patent sense) cannot be patented.

[87] This use of the term "abstract idea" corresponds to that of M.B. Nimmer, The Law of Ideas, 27 S. Cal. L. Rev. 118, 118 (1954).

Independent creation is likely in the case of reproductions, when the similarities may arise because both authors copied from the original. Independent creation of original works is also possible, since it is uncontroversial that similarly trained researchers working in the same field will often arrive independently at similar results. The near simultaneous invention of calculus by Liebniz and Newton is an example.[88] But when independently created, these works will undoubtedly be different in many details. It is only when described at a relatively high level of generality that they appear similar.

From this it follows that when a new work is described at a sufficiently high level of generality, similarity is not *prima facie* proof of copying, even though the work, described at that level of generality, is original to the author. This is because a defence of independent creation is plausible, just as the defence of independent copying from the original is plausible in the context of reproduction. It is also difficult to prove independent creation at this abstract level by showing lack of access since the gist of a work is more easily diffused than its details.

Just as in the case of reproductions, the fact that similarity is not sufficient proof of copying leads to a problem of proof. If a court attempts to determine on the facts whether the defendant copied the plaintiff's works, errors are inevitable. *Herbert Rosenthal Jewelry Corp. v. Kalpakian*,[89] in which the plaintiff complained that the defendant had copied its design for a jewelled bee pin, illustrates the problem. The Ninth Circuit held that

> there was substantial evidence to support the trial court's finding that defendants' pin was in fact an independent creation. Defendants testified to independent creation from identified sources other than plaintiff's pin. The evidence established defendants' standing as designers of fine jewelry and reflected that on earlier occasions they had designed jeweled pins in the form of living creatures other than bees, including spiders, dragonflies, and other insects, birds, turtles, and frogs. Any inference of copying based upon similar appearance lost much of its strength because both pins were lifelike representations of a natural creature."[90]

---

[88] The patent race problem is premised on the probability of independent invention: see the discussion *infra* n.000 and accompanying text. On independent discovery of new ideas, see generally Robert K. Merton, THE SOCIOLOGY OF SCIENCE: THEORETICAL AND EMPIRICAL INVESTIGATIONS, (1973).

[89] 446 F.2d 738 (9th Cir. 1971).

[90] *Id.*, 741.

But the court was not willing to decide the case on that basis: "Although this evidence would support a finding that defendants' bees were their own work rather than copied from plaintiff's, this resolution of the problem is not entirely satisfactory, particularly in view of the principle that copying need not be conscious, but 'may be the result of subconscious memory derived from hearing, seeing or reading the copyrighted work at some time in the past.'"[91] Since independent creation was likely on the facts, the Court simply could not decide with any reasonable certainty whether or not the defendant had copied from the plaintiff. In the result, the Court held that the even if there had been copying, what was taken was idea.

The decision of the Second Circuit in *Computer Assoc. Int'l, Inc. v. Altai, Inc.*[92] is another example. In the filtration step of the famous "Abstraction-Filtration-Comparison" analysis of copyright infringement of software, the court stated that "elements dictated by efficiency" were to be considered unprotectable. The main justification given was that in that case idea merged with expression. This is unhelpful when we are trying to decide where the line between the two is to be drawn. More helpfully, the Court also expressly noted that the utilitarian nature of programs and competition in the software market "give[s] rise to a problem of proof which merger helps to eliminate.... Since...there may be only a limited number of efficient implementations for any given program task, it is quite possible that multiple programmers, working independently, will design the identical method employed in the allegedly infringed work..... Under these circumstances, the fact that two programs contain the same efficient structure may as likely lead to an inference of independent creation as it does to one of copying."[93] The same argument applies to the other elements which the Court instructed must be filtered out, namely elements dictated by external factors, and elements taken from the public domain.

By analogy with reproductions, this suggests that "ideas" should be defined as those aspects of a work for which independent creation is likely but difficult to establish. This is

---

[91] *Id.*, citations omitted.

[92] 982 F.2d 693 (2d Cir. 1992).

[93] *Id.* at 708.

consistent with Learned Hand J.'s abstractions "test", but at the same time adds substance to it.[94] Of course this definition of ideas does not provide a mechanical test. "Likely" is not a precise term, and even if we did have a precise definition of likelihood it will not always be easy to tell at what level of generality independent creation is likely. We will return to these questions after fleshing out the basic argument.

Consider first the disadvantages, on a property rights analysis, of protecting ideas by copyright. The basic problem is the same as we identified in the context of reproductions: when independent creation cannot be established reliably, overprotection of the work is likely, with the attendant opportunism and consequent restriction on the use of new ideas. The analysis of the disadvantages of protecting ideas parallels almost exactly the problems which would arise if reproductions were protected. When independent creation is possible, the social value of the contribution of the first person to arrive at the idea is only the value of an earlier discovery date, which is of course less than the full value of the idea.[95] Thus the value of the earlier creation date for the idea corresponds to the reduced cost of access to the original provided by a reproduction, while the value of the idea itself corresponds to the value of the original. If property rights are well defined, the most that the first author of the idea can extract as a licence fee is the opportunity cost to the subsequent author of not having to independently create the idea, just as the most that a reproducer would be able to charge is the reduced cost of access. The private and

---

[94] The classic passage is found *Nichols* supra n.000 at 121: "Upon any work, and especially upon a play, a great number of patterns of increasing generality will fit equally well, as more and more of the incident is left out. The last may perhaps be no more than the most general statement of what the play is about, and at times might consist only of its title; but there is a point in this series of abstractions where they are no longer protected, since otherwise the playwright could prevent the use of his "ideas," to which, apart from their expression, his property is never extended." Judge Easterbrook has remarked that "Hand's insight is not a "test" at all. It is a clever way to pose the difficulties that require courts to avoid either extreme of the continuum of generality. It does little to help resolve a given case..." *Nash v. CBS, Inc.* (1990) 899 F.2d 1537, 1540 (7th Cir.). The property rights argument suggests that Judge Easterbrook's assessment is somewhat harsh. While Hand J.'s abstractions test is certainly vague as to where the line is to be drawn, it is not an entirely empty restatement of the problem, but points out that the more abstract the description of the work, the more likely it is that the description will fit an independently created work.

[95] The argument also applies when independent creation would have been likely but becomes impossible once other potential creators became aware of the first work. This would be the case if a creator who had almost completed a particular work came across the work of a rival who had arrived at a finished product. It might be impossible for the creator to forget the rival's work and take the last step independently, even though she would have developed it herself in due course. Even though the subsequent author did copy the work of earlier author, the value which the earlier author's work contributed to that of the subsequent author is only the value of the last few steps.

social benefits would then coincide. But if independent creation cannot be established, a subsequent independent author of an idea may have to licence from prior authors in order to avoid opportunistic lawsuits, just as a user of the original would have to licence from prior reproducers. This increases the private value of the idea, while the social value does not change. As with originals, some subsequent authors may choose not to use the idea because of the threat of opportunism. In the worst case no one will use the idea because any benefit would be more than lost to opportunism.[96]

While the analysis of the potential loss from granting protection to ideas parallels that of granting protection to reproductions, the analysis of the loss from denying protection is quite different. A major difference between ideas and reproductions is that an original work will exist even if protection is denied to reproductions, but it may be that some form of protection is required if the ideas are to be created at all. In the context of reproductions, the maximum loss from denying protection is only the increased cost of access to the original. This allowed us to be fairly confident in denying protection to reproductions. In the context of ideas, if protection is denied we may lose the ideas themselves entirely. Even rampant opportunism resulting from protection of ideas cannot be worse than that. Denying protection as a solution to the problem of poorly defined property rights is not nearly as appealing in the case of ideas as it is in the context of reproductions. A resolution of this dilemma is to grant patent protection instead of copyright.

II.C Potentially Patentable Ideas

Copyright protection will be denied to an original work when the court is of the view that even though the work may deserve protection, its nature is such that patent law is more appropriate.[97] This much is tautological. The difficult question is how to tell when patent rather

---

[96] Landes & Posner *supra* n.000 point out the difficulty of determining whether two authors have used the same ideas (at 350) and they also argue that protection of ideas would lead to rent-seeking (at 349), but they do not make the connection which is central to my analysis, namely that protection of ideas leads to rent-seeking if and only if a user might be found to have infringed an earlier work which she did not in fact copy from.

[97] *Baker v Selden* 101 US 99 (1880) is the leading case in point.

than copyright is appropriate.[98]

A fundamental difference between copyright and patent is that infringement of copyright requires similarity and copying, while infringement of patent requires only similarity. The property rights analysis suggest that this is crucial, because it means that patent provides better defined property rights than copyright when independent creation is likely but difficult to prove. This suggests that patent protection can provide an incentive to create ideas while avoiding the problem of opportunism which would arise under copyright protection.[99]

While this reasoning is a straightforward application of the property rights approach, it raises two puzzles. First, without a requirement of copying to add uncertainty to the scope of the right, a patent right will always be better defined than a copyright. Does the property rights analysis therefore imply that all works should be protected by patent? Secondly, copyright is undesirable when property rights are poorly defined because rent-seeking may dissipate the value of the idea. But the well-known patent race problem suggests that patent protection, by granting a *de jure* monopoly rather than a *de facto* one, will also dissipate the value of the invention through rent-seeking. Why then is patent protection any better than copyright?

The answer to the first question is relatively straightforward. Copyright is preferable to

---

[98] For convenience I will refer to patent protection and copyright protection, but I wish to abstract away from the fact that patent has a shorter term of protection than copyright and these terms should be understood to mean regimes in which the term of protection is the same. The question of the optimal term is in principle a separate policy question which depends on the point at which diminishing marginal returns to inventive effort set in and so requires empirical information on the "invention possibility function" for the category of works in question: see Nordhaus *supra* n.000, Scherer *supra* n.000. Ideally, for any given category of works we should decide whether to adopt a copyright or patent-style regime based on whether independent creation is possible and difficult to establish and then determine the optimal term of protection for that category. However, if such tailoring of the length of protection is not feasible (and perhaps not cost-effective considering the information requirements), so that our choices are, for instance, between a patent regime with a relatively short term and a copyright regime with a long term, it might be desirable to provide a category of works with the "wrong" form of protection if the term is much more appropriate. Further consideration of this issue is beyond the scope of this article.

[99] The argument closest to mine is that of David Friedman, Standards as Intellectual Property, 19 U Dayton L. Rev. 1109 (1994). He argues that "[w]here copying is easily accomplished and easily recognized and independent invention is unlikely, we would expect copyright law or something similar...[and] [w]here copying is expensive and hard to recognize and independent invention likely, we would expect to see something like patent protection" (at 1118). However, his is fundamentally a balancing argument. So, he argues that if copying is difficult residual incentives are more likely to be adequate; and copyright is more desirable if copying is easily recognized because this cost of enforcement is lowered, not because of the problem of opportunism. Friedman does point out the connection between independent invention and rent seeking, but does not discuss how patent law mitigates the problem as compared to copyright.

patent in two broad categories of cases: the first is when independent creation is possible, but it can be established reliably, and the second is when independent creation is inherently unlikely.

In the first case, when independent creation is possible but copying can be established reliably, copyright protection is superior to patent because patent protection gives rise to a rent-seeking problem and copyright does not. The rent-seeking problem, under the name of the patent race, is well recognized in patent law. With patent protection, the social value of an inventor's contribution is less than the private value of the patent: the first inventor can potentially extract the full value of the invention from any subsequent user, but, given that independent creation is possible, the social contribution of the first inventor is only the benefit of the earlier creation date.[100] This creates the potential for a race to be the first to patent, and this race can, in certain circumstances, dissipate the full appropriable value of the invention.[101] In contrast, under a regime of well-defined copyright protection, a rent-seeking "copyright race" is not possible. Because the second author has the option of independent creation, the maximum price which a user would pay to licence a reproduction is not the value of the original, but only the cost of access to the original. This maximum price is the same as the social benefit of using the reproduction. Since social and private value coincide, no rent-seeking problem arises.[102] Put another way, in contrast with the patent, a author who races to copyright a work first will not gain any licencing benefit thereby, since he will be undercut by later authors who waited until the cost of creation was lower before independently creating the work. So, when the rights are well defined, copyright is clearly superior to patent because it avoids wasteful rent-seeking.[103]

---

[100] In practice the inventor will not be able to extract the full value of the invention even during the term of the invention, as doing so would require perfect price discrimination. Also, there will be some unappropriable value remaining after the term of protection has expired. These factors depend on the pricing scheme and the term of protection, which do not depend on whether independent creation is a defence and so provide no reason for preferring a patent style of protection to copyright style protection.

[101] See the more detailed discussion *infra* at 000.

[102] This point is not always fully appreciated. For example Landes & Posner *supra* n.3 at 349 and separately at 350 invoke rent-seeking arguments in the context of copyright without noting that rent-seeking cannot arise when copyright is well defined: see *infra* footnote 000.

[103] There is some ambiguity. We cannot say that copyright is necessarily preferable to patent even in this context as there is some tendency to underinvestment in both regimes because the consumer surplus is not fully appropriable, so that some rent-seeking may be desirable: see e.g. J. Hirshliefer and J.G. Riley, The Economics of

The second case is when independent creation is not possible, as in the case of literal text of a work. In this case copyright does give rise to a *de facto* monopoly. Despite Hand J.'s dictum that an independent creator of Keats' Ode on a Grecian Urn would not infringe Keats' copyright,[104] a subsequent author would find it impossible to convince a court that he had not copied, at least subconsciously. But despite the monopoly, transaction costs cannot be higher than the value of the author's contribution, since the value of the expression and the value of the author's contribution will be the same. Similarly rent-seeking is not possible because try as they might *independent* creators will not arrive at the same expression. For the same reason, when independent creation is truly not possible, the rent-seeking problem does not arise under patent protection. The real advantage of copyright in this case is its lower administrative costs. There is no point to searching prior art, when independent creation is impossible in practice.

Why is Patent Protection for Ideas Desirable?

To this point we have seen that copyright protection for ideas is undesirable and copyright protection for expression is preferable to patent protection. The next question is why patent protection is desirable for ideas. How can replacing the *de facto* monopoly of a copyright in ideas with a *de jure* monopoly of patent mitigate the rent-seeking problem?

The most obvious suggestion is that patent law mitigates the problem of deadweight losses by providing protection to only to works which are a significant improvement. This is in contrast to copyright protection, which protects valueless works as well as valuable ones. It is true that a higher standard is desirable when independent creation cannot be established, since the problem of rent-seeking cannot be avoided entirely and so low value works should not be protected. But a higher standard is not sufficient since the problem of high licencing costs and

---

Uncertainty and Information – An Expository Survey, 17 J. Econ Lit. 1375 (1979). Also, there may be some duplication of effort under copyright since at least two independent creators are needed to create actual competition. Because of the sunk costs of creation the market cannot be considered fully contestable and a single creator charging prices just low enough to deter entry is possible scenario. This does not in itself give rise to rent-seeking because a monopolist creator will invest at a socially optimal rate, but there is a possibility of some wasteful investment in becoming the sole creator. In general the degree of inefficient effort under a copyright regime will depend on the detailed model of entry and competition and is beyond the scope of this article.

[104] *Sheldon v Metro-Goldwyn Pictures Corp* 81 F.2d 49, 54 (2nd Cir. 1936).

opportunism arises because of the difficulty of proving independent creation and so can occur with valuable works just as with trivial ones. There might at best be an indirect effect if risk-aversion caused rent-seekers to be less willing to over-invest in more valuable works.

Another suggestion might be that the patent filing system directly addresses one of the problems which arises from the likelihood of independent creation. By implementing a registration and filing system, patent law reduces the cost of searching for prior works which the second creator did not in fact use, thus making comprehensive *ex ante* licencing more feasible. But again this answer is not sufficient, since both the filing system and the higher standard of protection, while desirable when independent creation is possible, could equally be implemented in a regime in which independent creation is a defence.

A more persuasive answer focuses on the fundamental difference between copyright and patent, namely that in patent law independent creation is not a defence to an infringement action. While this feature of patent law is traditionally perceived as the source of patent-race problem, in fact it mitigates the problem of rent-seeking as compared to a copyright style regime when independent creation is likely.

As we have already noted, the potential for a patent race problem arises because the private reward from a patent is greater than the social contribution of the first inventor. But the fact that the first inventor will reap a reward which is greater than her contribution does not in itself create any inefficiency, as it simply represents a wealth transfer from subsequent users to the first inventor. The inefficiency arises in the potential for a race to capture the prize by being the first to invent. Since the second inventor gets nothing, if there is more than one potential inventor, each will increase R&D spending in order to bring forward their own expected invention date and thereby capture the patent prize. Each inventor will be willing to increase their investment until the expected value of the patent equals the total investment. Given diminishing marginal returns to the rate of R&D expenditure,[105] this will increase the total resources used to create the invention. In the aggregate, the individual inventors will be willing to spend an amount equal to the appropriable value of the patent prize. Except for marginal

[105] Returns to an increased rate of R&D spending will be diminishing both because there are likely to be diminishing returns to increasing the rate of spending on a particular project, and because the cost of undertaking any given project will decline with time as a result of general advances in the field.

inventions, this will be socially inefficient, since the socially optimum R&D expenditure maximizes the social value of the invention less the cost of invention. In the race to capture the patent prize, the inventors may dissipate the full appropriable value of the work over the patent period.[106] This may even lead to net losses, since the potential dissipation may be greater than the value of the first creator's contribution.[107]

But a race requires more than one entrant. If a single inventor knows that he is the only one working on a particular innovation, he will not dissipate the appropriable value of the patent in R&D, but rather will choose more leisurely investment schedule which will maximize his private return.[108] More generally, an inventor will attempt to maximize his expected return, and his expected return depends on his expectations regarding other potential inventors. If he is uncertain as to whether there are other inventors working on the same invention, he will reduce his R&D expenditure in order to increase his profits in the event that he is the first to invent. Put another way, the patent race will only dissipate the full appropriable value of the invention if the desirability of the invention is common knowledge, and the knowledge that it is common knowledge is also common knowledge.[109]

The advantage to patent protection then, is that property rights are granted when information regarding useful future inventions is limited. Thus Kitch argued that the patent system functions in part to reduce dissipation of social surplus in commercialization of an innovation by awarding an exclusive right of development.[110] An objection to Kitch's theory is

---

[106] See the review by Jennifer F. Reinganum, The Timing of Innovation: Research, Development, and Diffusion, Ch. 14 THE HANDBOOK OF INDUSTRIAL ORGANIZATION (1989).

[107] This is clearly true when the patent is granted for a known invention, such as playing cards, but it is also true if the invention would soon have been created even without granting patent protection. This justifies the obviousness requirement of patent law.

[108] He will not necessarily choose the socially optimal R&D expenditure, since the privately appropriable surplus and the social surplus are not the same.

[109] And, in principle, if knowledge that the knowledge is common knowledge is itself common knowledge, etc. For an interesting discussion of the problem of full rationality and common knowledge see John D. Geanakoplos, Common Knowledge, 6 J. Econ. Perspectives 53 (1992). These issues, though interesting, are not central to this point being made in the text.

[110] Edmund Kitch, The Nature and Function of the Patent System, 20 J. Law & Econ. 265 (1977).

that granting a 'prospect' patent simply pushes the rent-seeking into the race for the patent itself.[111] But if, as is posited by evolutionary theories of innovation, innovation proceeds myopically so that different firms have different information about and opportunities for innovation,[112] then not all opportunities are exploited immediately and rent-seeking will not result in complete dissipation of surplus. In a world of bounded rationality, Kitch's prospect theory is plausible.

Once we take imperfect information regarding feasible lines of research of into account in the rent-seeking problem, patent looks very different from poorly defined copyright. The rent-seeking problem would be significantly worse in a copyright system which protected works likely to be created independently, because claims of independent creation which are raised *after* the publication of the first creator's work would have to be taken seriously. There would then be an incentive for copiers who observe a valuable idea to claim to have created it independently in order to be able to extract licence fees from subsequent users of the idea.[113] In effect, a copyright regime would allow rent-seeking with the perfect information of hindsight. In contrast, patent protection eliminates the possibility of such *ex post* rent-seeking by rejecting claims by subsequent creators out of hand. Put another way, the patent system rejects the defence of independent creation and relies on bounded rationality in innovation to reduce rent-seeking, while the copyright system operates under perfect information regarding the value of the work and relies on the defence of independent creation to prevent rent-seeking.[114]

---

[111] D. G. McFetridge & D.A. Smith, Patents, Prospects and Economic Surplus: A Comment on Kitch, 23 J. Law & Econ. 197 (1980).

[112] See, most famously R. R. Nelson & S. G. Winter, AN EVOLUTIONARY THEORY OF ECONOMIC CHANGE (1982).

[113] It might be suggested that a court would never believe a claim of a subsequent creator, but if this were true even if independent creation were possible, then this would mean that the courts would in some cases reject the claim of someone who was in fact an independent creator of the work, and the system would be effectively a patent system.

[114] An additional advantage of the formal monopoly under a patent system is that a user need only licence from one prior creator for any given idea, which will directly reduce transaction costs. It might be suggested that granting a joint copyright to all simultaneous creators but not to subsequent creators would also solve the problem of rent-seeking with hindsight. But while such a regime would have fewer losses from rent-seeking than a copyright regime, it has the same rent-seeking losses as a patent regime and higher licencing costs. Thus it is inferior to a patent regime.

This assessment of the role of the patent monopoly also helps explain the rule that an invention is obvious if it is obvious to try and there is a reasonable expectation of success.[115] This rule is puzzling on the surface, particularly in the context in which it is most prominent, namely inventions using recombinant DNA technology to develop new drugs or other useful compounds. It may be obvious to attempt to produce some therapeutically useful human protein, such human insulin, using recombinant DNA technology, and success may be quite certain given enough time and money. At the same time, it will be much cheaper for a second team replicate the success of the first, once the essential information regarding the DNA sequence is known. The cost of initial development is very high, and the cost of replication is significantly lower, thus raising what appears to be a classic case for protection. In order to encourage research, we need to prevent the subsequent developer of the product from undercutting the party who did the essential research. Otherwise the admittedly and obviously useful product may never be developed at all. But the context in which the invention is obvious to try and probable to succeed is precisely when the threat of dissipation of the value of the intention through a patent race is the greatest. This is not completely compelling as a justification for the rule, since the patent race can only dissipate the appropriable surplus. The appropriable surplus is necessarily less than the social surplus both because the term of protection is limited and because the inventor cannot, without perfect price discrimination, which is impossible, appropriate all the value even during the term of protection. But even if the patent race problem does not strictly justify the rule in *O'Farrell*, it plausibly explains the source of the courts' concern.

Information Requirements

I have argued that the balancing analysis required to justify the idea/expression dichotomy is informationally so demanding as to be beyond the competence of the courts. The claim that the property rights approach is normatively and descriptively superior therefore requires a demonstration that it has more manageable information requirements. At this point in the discussion I will assume that similarity between works, as opposed to copying, can be reliably

[115] *In Re O'Farrell* 853 F.2d 894 (Fed Cir. 1988) is the leading case. English law has a rule which is similar, and, if anything, raises an even higher barrier for inventors: see *Genentech Inc's Patent* [1989] RPC 147 (C.A.) and *Biogen Inc v Medeva PLC* [1995] FSR 4 (C.A.).

determined, so that patent rights are always well defined. The issue of poorly defined patent rights is taken up in the next section.

First note that many of the issues which are important in a balancing analysis are irrelevant under the property rights approach. Denial of protection when property rights are poorly defined does not distort incentives for technological change which might result in a different balance. Since the analysis applies to on a case-by-case basis, if technology changes to allow better definition of the copyright, then it can and should be enforced when the technology is brought into effect. Similarly, there is no reason to change scope of copyright or patent protection as other incentives wax and wane, since those other incentives do not affect how well the property rights are defined. Nor need we be concerned about alternative forms of protection, since denial of copyright does not reflect a conclusion that the work should not be protected, but rather that copyright is not cost-effective. Even though alternative forms of protection, most obviously contract, may have higher transaction costs in some respects, they can achieve net gains by better *ex ante* definition of the property right.[116]

Now consider how the lines should be drawn between patent, copyright and unprotected works. Under the property rights analysis, the key question for the court in determining whether a work should be labelled an unprotected idea is the reliability of findings of infringement, and more particularly, the probability of false findings of infringement. In particular, the court should consider how likely it is that substantial similarity is the result of independent creation rather than copying and, if independent creation is possible, whether the ability to prove or disprove access will nonetheless allow reasonable determinations of copying to be made on the facts. Figure 1 summarizes the argument made so far, and can will serve to organize the following discussion. So, in the case of literal text ("Expression" in fig.1), independent creation is highly unlikely. This means that the court can reliably determine on the facts whether the defendant copied. No rent-seeking is possible, and the court can leave the assessment of value to the market. For this reason

[116] So the problem of preemption of contract law by copyright law is a red herring, and *ProCD v Zeidenberg* 86 F.3d 1447 (7th Cir. 1996), for example, was clearly correctly decided. Libott *supra* n.000 points out that the "format" for an television series which are purchased by a production company describes the show at a very high level of generality which would almost certainly be considered unprotected "idea" and protection is obtained by contract rather than copyright. The preemption argument implies, counter-intuitively, that this practice should be prohibited.

the likelihood of creation without protection, for which the value of the work is a loose proxy, is irrelevant. At the other extreme are ideas, which are works for which independent creation is likely, so that similarity does not establish copying.

The difficult question, of course, is how to draw the line between ideas and expression. As with the originality requirement, a form of balancing is required, because the value of the works whose dissemination will be restricted must be compared to the value of the works which would be denied protection. In the case of reproductions, we compared the value of the reproduction with the value of the original. Since the value of the original is often greater than the value of the reproduction (which is very little when the original is easily accessible) complete

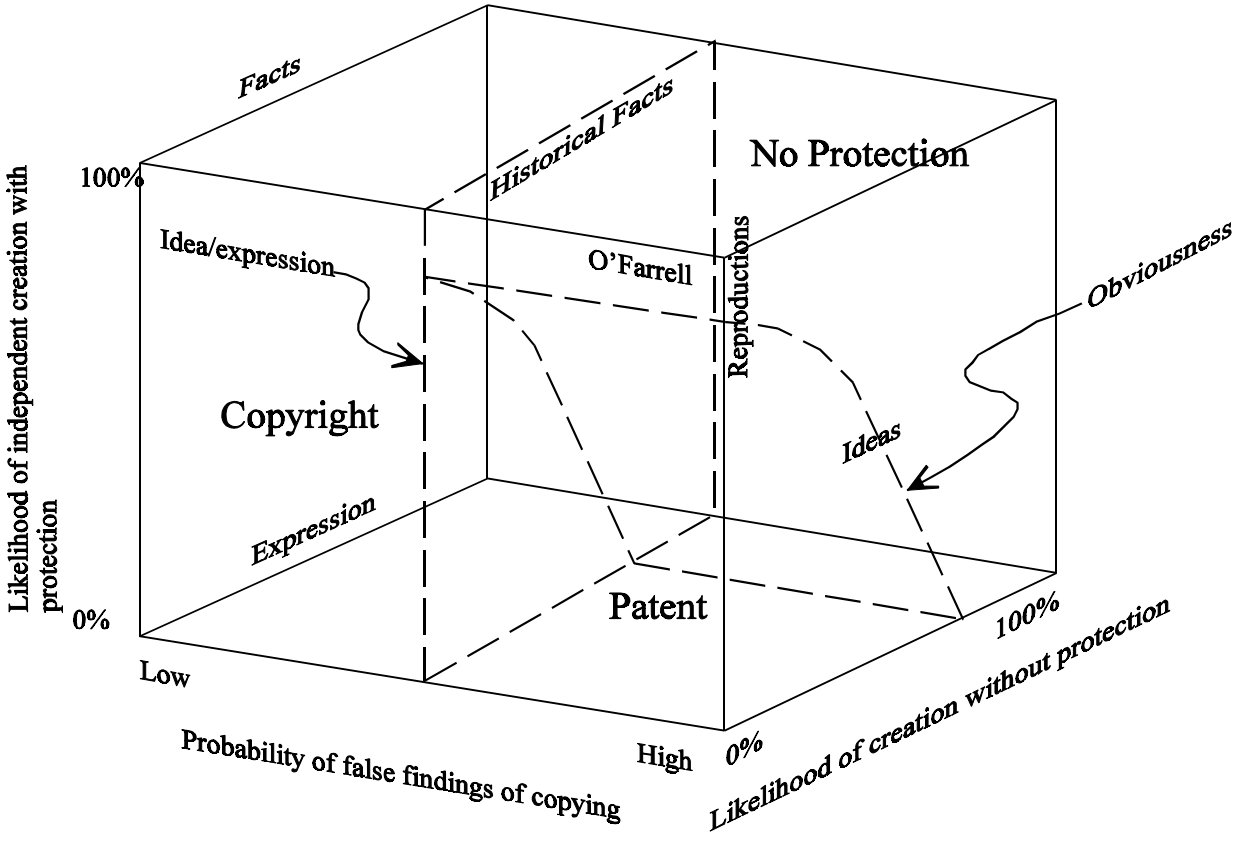
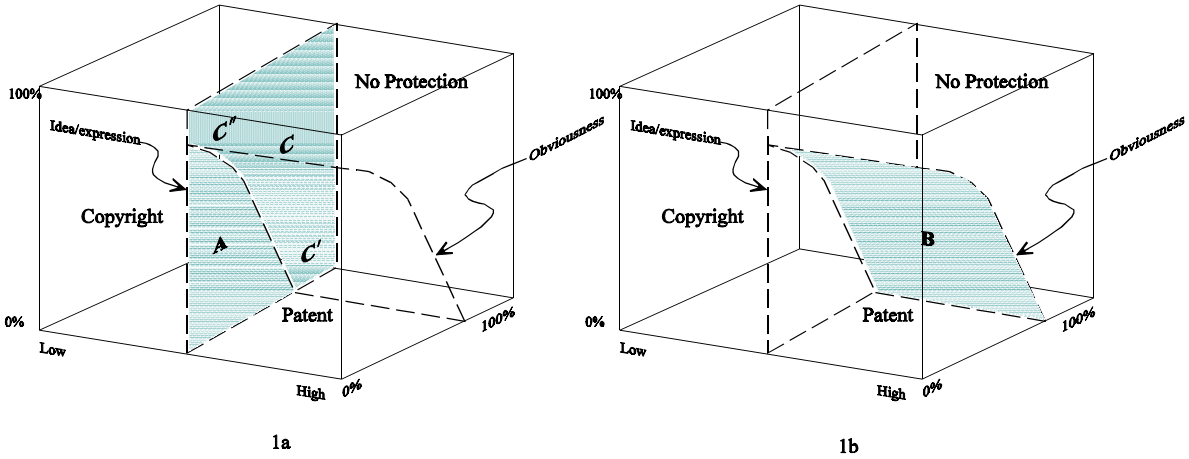


Figure 1



dissipation of the value of the original is not necessary in order to justify denying protection. In the case of ideas, we are concerned that granting protection to an idea will prevent use of the *same* independently created idea. However, granting protection may lead to an earlier creation date and this plays the same role as increased access to an original. At the extremes, if the idea is already widely known, then the benefit of an earlier creation date is zero and protection should be denied; but if the work would not be created at all without copyright protection, then protection should always be granted, since even complete dissipation would leave us indifferent between granting and denying protection. This implies that the court must assess how much earlier the work would be created if protection were granted in order to balance this against the dissipation of the value of the work. This is a difficult question as it requires a knowledge of the residual incentive structure in order to compare incentives with and without protection, as well as an estimate of the elasticity of supply, and this information is not readily available to a court in dealing with a specific case.

But, as we have seen, the choice is not just between copyright protection and no protection at all, since patent protection is an alternative. If it is very unlikely that the work will be created without any protection at all, but independent creation is likely and copying is difficult to establish if protection is granted, then patent protection is desirable. The question is not where the line should be drawn in denying protection to ideas, but where the line should be drawn between copyright and patent (shown as area A in Figure 1a.) When establishing independent

creation is moderately difficult we need to trade off the higher administrative costs of patent protection against the greater rent-seeking problems of copyright. This is certainly not easy, but errors should have only a second order effect on efficiency, since property rights are reasonably well defined under either regime and the question is one of choosing the regime with the lowest transaction and administrative costs. Drawing the boundary between copyright and patent is not easy, but errors are not crucial.

Not all works which are likely to be independently created are deserving of patent protection. When independent creation is likely, we have seen that rent-seeking will arise under either copyright or patent. This rent-seeking can actually destroy value, as when patent protection is granted to works which are not novel, such as playing cards. Thus patent protection is desirable if and only if the work will not be created sufficiently soon without protection. This is the problem of setting the obviousness threshold in patent law. This is an irreducibly difficult balancing problem which is not made any easier by the property rights. However, the property rights analysis does add some insight. Patent protection is beneficial because it induces an earlier creation date for new works. But as the probability of independent creation given patent protection rises, the rent-seeking problem increases. For this reason, as the probability of independent creation given patent protection rises, we should require more benefit in terms of an earlier creation date. Conversely, we should be more reluctant to grant protection, even though the work may not be created without protection. This is why the "obviousness" boundary (area B in Figure 1b) slopes downwards.

The final question is how to draw the line between works copyright protected works of low value, and works which are not sufficiently valuable to be protected by patent. As noted, when independent creation is very unlikely, so that there is no difficulty in establishing independent creation, the courts need not be concerned with the value of the work. But if the work does not satisfy the obviousness standard of patent, and the defence of independent creation is sufficiently difficult to establish, both copyright and patent protection should be denied. This is the question of where boundary C in Figure 1a should be established. If the non-obviousness standard in patent law is not set very high, then this issue is largely absorbed into the two previous questions. (Note that area A and area C are contiguous.)

To the extent that it is does pose an independent question, drawing the exact line is a relatively difficult balancing problem, but as in the case of drawing the line between copyright and patent, the consequences of error are small. Errors in denying protection to works which are likely to be created even without protection (in the neighbourhood of C′) are evidently of little importance, because granting protection would only marginally advance the date at which they are created. Granting protection in such cases is more problematic, because the rent-seeking problem can lead to a dissipation of value which is greater than the benefit of an earlier creation date. This suggests that the courts should err on the side of denying protection when the works in question are of low value. An example of this is denial of protection to reproductions when the original is readily available.

When the work is not likely to be created without protection, but it is likely to be created if some protection is granted (area C″), then the benefit of granting either copyright or patent is minimal because of the rent-seeking problem. Again, it is difficult to determine where the line should optimally be drawn, but errors will have only a second order effect. Also, we have seen that there is an alternative to simply denying or granting protection. When the work is unlikely to be created without protection, but the nature of the work is such that independent creation is difficult to establish in general, the court can avoid the rent-seeking problem while providing at least some incentives for creation by granting copyright protection when the court is confident on the facts of a particular case that there was copying.[117] Note that this suggestion is inconsistent with a balancing analysis, in which the point of denying protection is to allow copying without the necessity of incurring the associated transaction costs.

II.D Abstract Ideas

A final category of ideas to be considered are those which are protectable by neither

[117] For an example of the court protecting relatively abstract elements when copying is clearly established on the facts, see *Metro-Goldwyn-Mayer, Inc., v. American Honda Motor Co.*, Inc., 900 F. Supp. 1287; 1995 U.S. Dist. LEXIS 19169 (C.D. Cal. 1995), which found infringement of James Bond films by the defendant's car chase commercial which was code named "James Bob". See also the discussion of reproductions *supra* n.000 and accompanying text. Of course it could be argued that these cases were wrongly decided or that these elements are indeed protectable expression or sufficiently 'original' respectively. Any reasonable number of examples would not be conclusive and ultimately the reader must judge for him or herself the descriptive accuracy of the claim that the courts are more likely to find copying of protectable expression when copying is established on the facts.

copyright nor patent, namely abstract and literary ideas. Even if we accept that such ideas should not be copyrighted because of the likelihood of independent creation, why is patent law not appropriate? The answer is that property rights are poorly defined even under a patent regime when similarity between works cannot be reliably established.

Literary ideas are only manifested via expression which must be interpreted by a judge to extract the idea. The higher up the chain of abstraction from expression to idea, the greater the proliferation of ideas which are plausibly attached to a particular expression. As a result, after a second idea is developed, it may well be possible to see the same idea in an earlier work, even if that idea had not been seen there previously. When an English scholar gives a new reading to *Hamlet*, has she created a new idea, or simply discovered one which had always been there, albeit undetected? Conversely, because of the difficulty of objectively distinguishing abstract ideas, an earlier author might be able to make a plausible claim that his idea is the same as that in a second (more successful) work even though, on the surface, the two seemed so different that the second author would never have thought to seek a licence from the first.[118] In this case opportunism could not be avoided even under a patent system.

The rule that "disembodied mathematical idea[s]" or "abstract ideas" are not patentable is justifiable as a response to the problem that patenting of abstract ideas would result in poorly defined property rights. Patent rights can never be perfectly well defined, since after all, the claims which nominally serve to define the scope of the monopoly are literary text which must be interpreted by the court. The problem of *ex post* reinterpretation of a patent means that the scope of licencing is never entirely clear *ex ante*. But the requirement that a patentable work must be described in a specific embodiment reduces the uncertainty over the scope of the claim. For example, if a mathematical formula were patentable in itself, a subsequent creator would have to determine whether his invention could possibly be described as being an embodiment of that

[118] Landes and Posner, *supra* note 000 at 350, make the point that enforcing copyright in ideas would be difficult because a second author would attempt to make his idea seem different by expressing it differently. This is a related but distinct point. Landes and Posner are raising the problem of underprotection, that is, the failure to find infringement when there was in fact copying, which is relevant to a balancing analysis as it reduces the incentives to create generated by copyright protection. As we have seen, under the property rights approach this problem is much less important than that of overprotection, i.e. the problem of finding infringement when there was no copying, which leads to the potentially greater costs of opportunism.

formula. Since the same phenomenon may often be described from different physical or mathematical perspectives, this may not be obvious. Indeed, it may only be after considerable time has passed that the fundamental similarity of different phenomena is recognized. On the other hand, if it is only a specific application which can be claimed then the subsequent inventor need only determine whether the claimed application also describes the application of his invention.

The precise scope of the prohibition of patenting ideas, particularly algorithms, is still the subject of controversy, with an evident division between the Federal Circuit and the Supreme Court. The property rights analysis I have outlined strongly supports the Federal Circuit court's holding in *In re Alappat*[119] that the crucial point is whether or not the subject matter of the patent application is "abstract" as opposed to specifically embodied. It is inconsistent with the Supreme Court's position that a work should be unpatentable if all that is novel is the idea, even though the idea is embodied in a specific application.[120]

The judgment as to whether the patent claims an invention which is sufficiently concrete that it can be distinguished from other inventions is an issue which is clearly within the competence of the courts to determine on the basis of the available information.

II.E Critique of Functionality as a Criterion for Demarcating Patent from Copyright

The property rights analysis of the boundary between copyright and patent is usefully contrasted with the common intuition, most clearly developed by Professor Karjala, that "[p]atent law protects creative but functional invention; copyright law protects creative but nonfunctional

---

[119] 33 F.3d 1526 (1994 Fed. Cir.)

[120] See *Gottschalk v Benson*, 409 U.S. 63 (1972); *Parker v Flook* 4376 U.S. 584 (1978); *Diamond v. Diehr* (1981) 450 U.S. 175,; and see the dissent in Archer J. in *Alappat* pointing out the inconsistency between the majority opinion and the approach taken by the Supreme Court. A common objection to patenting of a novel non-obvious algorithms, is that this can grant an effective monopoly over the use of algorithm, particularly if its uses are restricted: see e.g. *Benson* at 71-72. This is a version of an optimal balancing argument which is unpersuasive without an assessment of the offsetting reduction in incentives to create. Of course, even if the work is uncopyrightable as being an idea but is patentable subject matter, it will not necessarily meet the patent law tests for patentability, in particular non-obviousness. So the specific result in *Benson*, for example, is defensible because the algorithm in question was trivially obvious, although for reasons which are unclear this argument was not raised.

authorship."[121] The distinction between functional and non-functional works is intuitively appealing and has some descriptive validity. However, it is not sufficiently robust to sustain a formalist argument that copyright should protect only non-functional works because such a rule provides the best descriptive fit with existing law. The very similar view that patent protects items which are used in industry and technology, while copyright protects cultural items has been clearly rejected by the Supreme Court.[122] To accommodate this Karjala distinguishes between "functional" and "useful" works, saying "a work is "functional" if it performs some utilitarian task *other than* to inform, entertain, or portray an appearance to human beings."[123] The exclusion of works that "inform, entertain or portray an appearance to human beings" is fairly blatant gerrymandering, designed to exclude the functional aspects of uncontroversially copyrightable works such as photographs and maps. It implicitly acknowledges that informing, entertaining or portraying an appearance to human beings is "functional" in an intuitive sense of the word and there is no principled reason for excluding such works other than to tailor the definition to the scope of undisputed copyright protection. It is also inconsistent with the use of the term "functional" in trade mark law, which does include entertaining and depicting a pleasing image as functional.[124] And even with this exclusion, works such as novels may also be functional in a more restrictive sense, for instance when a boring book serves as a substitute for sleeping pills. And Karjala admits that program code, which is undoubtedly protected, is functional.[125]

Thus, as Karjala forthrightly acknowledges, the main force of the argument that patent

---

[121] Dennis S. Karjala, Copyright Protection of Computer Documents, Reverse Engineering, and Professor Miller, 19 Dayton L. Rev. 975, 976-77 (1994) ("Reply to Professor Miller"); and see Dennis S. Karjala, Copyright, Computer Software and the New Protectionism, 28 Jurimetrics J. 33 (1987); Dennis S. Karjala & Peter S. Menell, Brief Amicus Curiae: Applying Fundamental Copyright Principles to *Lotus Development Corp. v. Borland International, Inc.* 10 High Tech. L.J. 177 (1995).

[122] *Mazer v. Stein*, 347 U.S. 201, 74 S. Ct. 460 (1954)

[123] Reply to Professor Miller, *id.* at 977, emphasis added.

[124] See e.g. *Pagliero v. Wallace China Co.*, 198 F.2d 339 (9th Cir. 1952), holding a decorative design on china to be functional; *International Order of Job's Daughters v Lindeburg & Co.* 633 F.2d 912 (9th Cir. 1980) holding that aesthetic components of the jewelry are functional.

[125] Reply to Professor Miller, *supra* n.000 at 978. On the copyrightability of code, see *Apple Computer, Inc. v Franklin Computer Corporation*, 714 F.2d 1240, 1253 (3rd Cir. 1983) cert. denied 464 U.S. 1033 (1984).

law is preferred for protecting functional works must be normative.[126] He begins his normative argument by noting that because technology proceeds incrementally patent protection denies protection to minor improvements in order to allow others to build on earlier works. This is the basic point that because IP protection entails deadweight losses it should be limited. The difficult problem is to explain the low standard for protection in copyright. On this point Karjala argues that in the non-functional field "variety is the spice of both legal and real life" and so "[t]he social utility of allowing subsequent authors to make minor variations on a copyright-protected novel is minimal."[127] This is simply not true. Minor variations on copyright protected works, such as sequels to popular movies, genre novels, and derivative works generally, are legion and their social utility, at least when measured in terms of dollar sales, is very high. Further, Karjala's argument is premised on the notion that copyrighting works with little originality will encourage authors to create new works, presumably in order to avoid the need to pay a licence fee. This is not correct either. When licencing costs are low, broad protection does not decrease the supply of minor variations. This is why derivative works are in fact so common even though the right to produce them must be licenced from the original author.[128] Thus drawing the line between copyright and patent on the basis of functionality is neither descriptively nor normatively satisfactory.

While functionality does not accurately distinguish between patentable and copyrightable works, there is nonetheless some division of labor between copyright and patent which is

---

[126] *Id.* at 982.

[127] *Id.* at 979. Chafee, Reflections on the Law of Copyright, *supra* n.000 at 512, makes a very similar argument.

[128] Two other arguments regarding the relative scope of copyright and patent should be considered. Wiley *supra* n.000 hesitantly argues that "the creation of new systems and processes is the sort of innovation that benefits from initial library research" which contrasts with "the writing of music, plays, and fiction – where judges lack the confidence to insist that creators must begin with library research" at 166, 167. His argument on this point is full of qualifiers and his reservations are well warranted. Copyright protection does not preclude authors from beginning with library research, but rather encourages them to do an optimal amount of library research: authors have an incentive to do library research so long as doing so is a less costly way of developing their work than is independent creation. When independent recreation is not an infringement the need to licence works which are found as a result of library research does not affect the incentives to do such research, as the licence fee will only reflect the savings of effort. This means that even if systems are in fact better developed by beginning with library research, there is no reason for the courts to require such research. Further, there is simply no reason to believe that library research is a more efficient way to develop systems than it is to create novels or advertising copy; the argument is purely *ad hoc*.

imperfectly captured by the functional/artistic distinction. The property rights argument explains this rough generalization as a by-product of using patent law to protect works which are likely to be independently discovered. The value of a functional work generally lies in the function it serves rather than the precise means which it uses. This means that protection is needed at a relatively high level of abstraction if it is to be useful: Baker would not have been satisfied if the Court had given him protection for his forms but not his system of bookkeeping, since the defendant could and in fact did use a slightly different set of forms to accomplish the same end.[129] To the extent that the value in a functional work lies at a level at which independent recreation is likely, and so copyright protection should be denied, it is not surprising that patent protection should be necessary. On the other hand, since considerable value in a non-functional work lies at the level of expression, that is, in details which are unlikely to be independently recreated, copyright protection is more likely to be adequate and will be preferred because the term is longer and it is more easily obtained. Similarly, when the value of a useful work is found at a low level of generality so that copying is unlikely or easily disproved, for example in the case of maps or the shape of a decorative statuette/lamp, copyright protection can be and is sought.[130] So even though patent more usually protects functional works and copyright non-functional works, this distinction is not a sound normative basis for deciding which regime of protection is appropriate, but is simply an artifact of allocating to patent law protection of works for which independent creation is likely.

II.F Summary

The property rights approach to the limits to copyright which is described above requires trade-offs to be made between denying protection to ensure clear property right and ensuring incentives to create. These trade-offs are not trivial. But I do not argue that the property rights analysis allows for a simple inquiry and a clear result in all cases; any analysis of the idea/expression dichotomy which arrived at a simple and certain test could not lay claim to being remotely descriptive of the law. Rather I suggest that the property rights approach allows an

---

[129] *Baker v. Selden* 101 U.S. 99, 100 (1879).

[130] See *Mazer v Stein*, 347 U.S. 201, 74 S. Ct. 460 (1954)

analysis which, unlike the balancing approach, is generally reasonably within the competence of the courts.

A key feature which distinguishes the balancing approach from the property rights approach is that the balancing approach trades off incentives to create a work with restrictions on the dissemination of that same work, while the property rights approach trades off incentives to create a work with restrictions on the dissemination of *other* works. In the context of reproductions, we were not concerned that granting copyright in the reproduction would unduly restrict access to the reproduction itself; rather, we were concerned that it would restrict access to the original. Similarly, on the property rights approach we are not concerned that granting copyright in an idea to an author who has developed it will prevent others from benefiting from that author's creation. Rather we are concerned that other authors, who have created the same idea independently, might be prevented from using their own idea, by the fear of an opportunistic action by the earlier author.

Judicial pronouncements as to the goals of copyright are often ambiguous in respect of this distinction. For example, the Supreme Court in *Twentieth Century Music Corp. v. Aiken* stated that "The limited scope of the copyright holder's statutory monopoly. . . reflects a balance of competing claims upon the public interest: Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts."[131] This is as consistent with the property rights approach, as it is with the balancing approach, since poorly defined copyright will indeed reduce the broad public availability of new works. While most judicial pronouncements are similarly ambiguous, there are some cases which clearly adopt a balancing approach,[132] while the property rights

---

[131] 422 U.S. 151, 156, 95 S. Ct. 2040, 2043-4 (1975).

[132] The Supreme Court's decision in *Feist* being a notable example. The Court's statement, quoted *supra* n.000, that granting a monopoly in facts would preclude subsequent authors from relying on those facts, clearly evidences a concern for restrictions on the dissemination of the very work for which protection was being sought. Similarly, *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429, (1984) (emphasis added): "As the text of the Constitution makes plain, it is Congress that has been assigned the task of defining the scope of the limited monopoly that should be granted to authors or to inventors in order to give the public appropriate access *to their work* product. Because this task involves a difficult balance between the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand, our patent and copyright statutes have been

approach has less explicit support.[133] That the property rights approach is not supported in express judicial pronouncements is of course not a strong critique of the approach, since the point of academic analysis is to find the substance behind the judicial rhetoric. And precision is not important in the easy cases, since either formulation can serve as a prop for intuition.[134] The correct analysis is much more important in novel or borderline cases. So, for example, one of the clearest statements of the balancing approach is found in the Supreme Court's decision in *Feist*, and in the next Part I argue that this case was incorrectly decided.

III FACTS, FAIR USE AND DERIVATIVE WORKS

III.A Facts

In *Feist Publications Inc v Rural Telephone Service Co. Inc.*[135] the Supreme Court held that there can be no copyright in facts. The Court justified its decision on two main grounds. One, we which have noted previously, is that ideas are not protected in order to allow subsequent authors to build freely upon them, and facts like ideas are used as building blocks for future

---

amended repeatedly." *Herbert Rosenthal Jewelry Corp. v. Kalpakian* 446 F.2d 738, 742 (9th Cir. 1971) is an interesting example, as the Court made a clear statement of the balancing approach's concern for the owner's monopoly in their own work, stating that "The guiding consideration in drawing the line is the preservation of the balance between competition and protection reflected in the patent and copyright laws. What is basically at stake is the extent of the copyright owner's monopoly -- from how large an area of activity did Congress intend to allow the copyright owner to exclude others?" But the substance of the decision is nonetheless entirely consistent with the property rights approach: see the discussion *supra* n.000 and accompanying text. *Sayre v Moore supra* n.000 is another cases which clearly applies a balancing approach, saying that a servile imitation is an infringement, but taking facts from the same work is permissible. Landes and Posner *supra* n.000 are also clearly concerned with access to and incentives to create the a given work: see generally the introductory discussion and esp at 326: "Copyright protection...trades off the costs of limiting access to a work against the benefits of providing incentives to create the work in the first place." Macaulay's analysis was also expressly one of balancing : "...the evil effects of the monopoly are proportioned to the length of its duration. But the good effects for the sake of which we bear with the evil effects are by no means proportioned to the length of its duration. A monopoly of sixty years produces twice as much evil as a monopoly of thirty years, and thrice as much as a monopoly of twenty years. But it is by no means the fact that a posthumous monopoly of sixty years gives to an author thrice as much pleasure and thrice as strong a motive as a posthumous monopoly of twenty years. On the contrary, the difference is so small as to be hardly perceptible." See *supra* n.000 at 180-81.

[133] I have argued supra n.000 that *Herbert Rosenthal Jewelry Corp. v. Kalpakian* and *Computer Associates v Altai* are clear but implicit applications of the property rights approach.

[134] See generally the cases cited *supra* n.000.

[135] 111 S. Ct. 1282 (1991).

works, so facts, like ideas, should not be protected. While the minor premise, that facts are like ideas in that both can be used as building blocks, is correct, I have argued that the major premise, that ideas are denied protection because they are building blocks is not correct, so the argument is not persuasive.

I will not repeat my general criticism of the building block argument[136] but we should consider how it applies on the facts of *Feist*. First, as I have noted, the court's main premise, that granting copyright protection to facts would "absolutely preclude" subsequent authors from relying on those facts, is simply wrong. Copyright protection would not prevent subsequent authors from using the facts, it would only require them to pay for the privilege. It might be suggested that the facts in *Feist* show that protection for information will indeed result in restricted dissemination because the defendant had attempted to licence the information but negotiations had broken down.[137] There are two problems with this suggestion. First, even if the bargaining breakdown in *Feist* occurred because of a bilateral monopoly, this does not justify denying protection to facts since bilateral monopoly is always a possibility in respect of any aspect of a work which is protected by copyright, or any other form of protection for that matter. Bargaining breakdown is as likely to occur in licencing expression as licencing of facts and it therefore does not provide any reason for distinguishing facts from expression. Second, it seems likely that the bargaining broke down, not because of bilateral monopoly, but because of lack of clarity in the law. Despite the Supreme Court's assertion that facts and ideas have always stood on the same footing, and that facts had never been copyrightable, the pre-Feist law regarding copyright in facts was not at all clear.[138] This is illustrated by the fact that Feist saw fit to

---

[136] *Supra* n.000 and accompanying text.

[137] Feist wished to assemble an area-wide telephone directory which would provide white and yellow pages for an area served by eleven different local telephone companies. Feist's directory would be distributed free of charge, financed by advertising. While the competition from Feist would lower the advertising revenues from the local telephones companies' own yellow pages, presumably a single consolidated directory is more valuable than eleven local directories, so that the total revenues from advertising less costs would be greater for the consolidated directory. (If this were not the case, there would be nothing inefficient about bargaining breakdown.) Of the eleven local companies involved, only Rural refused to license its white pages listing to Feist.

[138] The pre-*Feist* case-law was inconsistent with a significant "sweat of the brow" school affording broad protection, and *Feist* is generally considered to have affirmed one of the more radical poles of the spectrum of positions: see R. A. Gorman, The *Feist* Case: Reflections on a Pathbreaking Copyright Decision, 18 Rutgers

negotiate over, and in ten of eleven instances, to pay for information which, according to the Court's ultimate decision, could have been freely appropriated. Objective uncertainty in the law implies a greater likelihood in large differences between subjective valuations of the property right which Rural could assign (or equivalently, in subjective estimates of the probability of success in litigation), and a concomitant increase in the probability of bargaining breakdown.[139]

The Court's decision does seem to have been based primarily on the building block theory, and not on a balancing approach more generally. I say this because intuition suggests that the Court's decision in *Feist* is not defensible on a balancing argument, although of course this conclusion is tentative in view of the general difficulty of balancing arguments. Factual works are highly susceptible to free-riding and however plausible the suggestion that the value of the expression alone will allow the author to recover her costs in the case of literary works such as novels, it is much less plausible in the case of factual works, as a large part of the value of a factual work will be independent of the expression of the facts it contains. The presentation of the facts will of course contribute to the value of the work, but in contrast to novels and plays, different expressions of the same facts may be almost perfect substitutes for the original work.[140]

The second main argument advanced by the Court is that facts are not copyrightable because they are not original: "...facts do not owe their origin to an act of authorship. The distinction is one between creation and discovery..."[141] The Court apparently viewed this as a purely textual point, as no policy distinction between creation and discovery was offered.

Computer and Tech. L.J. 731, 25 I.P.Law Rev. 355 (1992); J. Ginsburg, No "Sweat?" Copyright Protection and Other Protection of Works of Information After *Feist v. Rural Telephone*, 92 Colum. L.Rev. 338 (1992); D.B. Wolf, Is There Any Copyright Protection for Maps After *Feist*? 39 J. Copyright Soc'y 224 (1992); L. Raskind, Assessing the Impact of *Feist*, 17 U.Dayton L.Rev. 330 (1992).

[139] See E. Hoffman and M.L. Spitzer, The Coase Theorem: Some Experimental Tests, 25 J.Law & Econ. 73 (1982).

[140] It is true that the listing information in *Feist* will continue to be produced even without copyright protection, since Rural's local franchise monopoly required it to publish a directory, and because denial of protection reduces licencing costs the decision is arguably justifiable on its specific facts. However, the Supreme Court clearly did not confine its decision to the facts, but intended it to apply to factual works generally. The relevant question is therefore not on whether copyright protection was necessary to produce the information at issue in *Feist*, but whether in general there is a sufficient incentive to produce factual works even without copyright protection.

[141] *Supra* n.000 at 1288.

However, it does hint at a more subtle argument: to the extent that facts are objective representations of the world, independent creation of a given factual work is possible. On the property rights analysis presented above the likelihood of independent creation is a far more relevant analogy between facts and ideas than the observation that both can be used as building blocks for future works.

But while facts are unlike expression in that independent creation is possible, they are unlike ideas in that such independent creation is nonetheless generally easily provable since creation of factual works requires a concerted effort which will provide plentiful evidence. Alternatively, the common practice of inserting minor errors in the work (as was done by the plaintiff in *Feist*) allows copying from the plaintiff's work to be established with certainty, and on the property rights analysis protection should therefore be granted. So long as independent creation can be reliably established, the problem of deadweight losses is no greater for facts than for the protection of original expression. The problem of "abstractness" that is, difficulty in proving similarity, is evidently not a concern with respect to facts, for which similarities and differences are at least as objectively determinable as for literal expression. So, while facts are not original in the sense that independent creation is possible, the property rights analysis implies that originality is simply a proxy for ease of proving copying. That similarity may arise because of independent creation does not justify denying protection if independent creation can be readily established through direct evidence despite similarity and access to the first work.[142] Thus *Feist* was undoubtedly wrong in saying that copyright-style protection should be denied to facts.[143]

The argument for denying protection to news and historical facts or theories is much stronger. Once a historical theory, e.g. that it was not Dilinger but a double who was shot by the

---

[142] So, in Figure 1, I have shown facts as having a high probability of independent creation, but low probability of a false finding of infringement.

[143] Consistently with evolutionary/structural arguments, legislative reform of the law in this respect seems inevitable. The E.U. has passed a Databases Directive, Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (see http://www2.echo.lu/legal/en/ipr/database/database.html) which is binding on member states, and which has been implemented, for example in the U.K. by the *Copyright and Rights in Databases Regulations 1997*, SI 1997 No 3032. In the United States lobbying is active, resulting most recently in H.R.354 *Collections of Information Antipiracy Act* (106th Congress).

FBI outside the Biograph Theatre in 1934[144], is revealed, it is not possible to independently re-create it. But if the theory is true or at least plausibly supported by evidence, then independent creation would otherwise have been possible. Two independent researchers are more likely to reconstruct the same story when attempting to discover historical truth than when attempting to write historical fiction. Thus historical theories are analogous to 'patentable ideas' discussed above.[145]

News presents a somewhat more difficult case. While the Court in *International News Service v Associated Press*[146] stated that there was no copyright in news, it did provide effective protection through the doctrine of misappropriation. In effect the Court established a *sui generis* regime of protection for news with a much shorter term than copyright. The decision lends itself to being read as an attempt at optimal balancing of incentives to create and restrictions on dissemination; protection is needed to recoup the cost of newsgathering, but if the value of news is very high initially and declines steeply thereafter, a form of protection with a short term allows the creator to recover most of the value of the news while saving transaction costs for low value users in the longer term. This balancing exercise does require an estimate of the shape of the demand curve for news, but in this case the rough estimate is perhaps a plausible one. The difficulty with this argument is that if it is true that there are many low value users in the long run, so that transaction costs are high, news agencies would not be able to collect significant revenues from these users even with protection, precisely because of the high transaction costs. In a competitive newsgathering market, we should therefore expect news agencies to release the basic facts into the public domain in order to make their product more attractive to the higher value users. As always, the plausibility of the balancing argument turns on specific aspects of the market for the good in question.

The *Associated Press* decision has a straightforward justification on the property rights approach. It may be easy to establish that the defendant copied from the complainant's news

---

[144] See *Nash v. CBS, Inc.*, 899 F.2d 1537; 1990 U.S. App. Lexis 6346 (7th Cir. 1990).

[145] See Figure 1.

[146] 248 U.S. 215, 39 S. Ct. 68 (1918).

bulletins in the first hours in which the news appears, but at a later time the defence that the news was gathered independently becomes much more plausible. On this view, the scope of protection was narrowed, not to avoid transaction costs associated with licencing, but in order to restrict protection to that period in which copying can be reliably established.

III.B Fair Use

Previous economic analysis have justified fair use largely as a means of increasing dissemination of a work when transaction costs of licencing are higher than the benefit to a subsequent author of using the work.[147] This necessarily requires a difficult balancing analysis.[148] The property rights approach does not challenge this basic understanding of fair use. However, the balancing approach is incomplete and can be usefully extended by the property rights analysis.

The difficulty with the balancing argument is that it assumes that when transaction costs are higher than the value of the work to a user, so that licencing is not cost-effective, the user will not be able to use the work. This is not correct. If a consumer chooses not to licence a work of intellectual property because the transaction costs exceed the value of the work to him, he will nonetheless rationally choose to use the work without a licence unless the expected loss from an infringement action by the author of the unlicensed work is greater than the opportunity cost of not using the work. So, if the cost of a lawsuit is at least as great as the costs of *ex ante* bargaining and the maximum licence fee which the author can extract *ex post* is no more than could have been extracted *ex ante*, in any case in which the consumer was priced out of the market *ex ante* the cost of a lawsuit will be greater than the potential licence fee and so the threat to sue will not be credible. Anticipating this, the consumer will choose to use the work without licencing and dissemination will not be restricted, notwithstanding that licencing costs are higher than the work's value to the user.

---

[147] Wendy J. Gordon, Fair Use as Market Failure: a Structural and Economic Analysis of the Betamax Case and its Predecessors, 82 Colum. L. Rev. 1600 (1982) at 1629; Landes and Posner *supra* n.000 at 357.

[148] Because fair use is applied on a case-by-case basis, balancing in the context of fair use is somewhat less objectionable than in respect of doctrines such as the idea/expression dichotomy, although accurate balancing is still very difficult. See generally the excellent analysis of Fisher, Fair Use Doctrine *supra* n.000.

We may safely assume that *ex ante* licencing costs are lower than *ex post* litigation costs, but it is not likely to be generally true that the maximum *ex post* licence fee will be no greater than the maximum *ex ante* fee. In particular, if sunk costs are incurred in the use of the work, then it may be rational for a creator to sue *ex post* even though it would not have been rational to licence *ex ante*. An adverse selection problem also arises. If there is significant uncertainty in the outcome of a project, the expected profit might be such that licencing the plaintiff's work would make the work infeasible, yet a prior creator would find it worthwhile to sue in respect of the most successful projects. In either case it is high *ex ante* licencing costs relative to the value of the work to the user combined with the threat of opportunism which leads to restricted dissemination.[149]

Since consideration of the threat of opportunism adds to, rather than replacing, the need to compare *ex ante* transaction costs with the value of the work to the user, it does not generally reduce the information requirements of the fair use doctrine. The comparison of *ex ante* transaction costs with the value to the user is an informationally demanding balancing analysis and opportunism itself is not always easy to identify. High individualized transaction costs which prevent price discrimination arise in considerable part because the user's surplus may be private knowledge. This means that when a creator is considering launching an infringement action she may not be able to accurately gauge the benefits and "opportunistic" suits may result simply from the creator's efforts to minimize administrative costs of policing its pricing policy in the face of a knowledge asymmetry. Because the relevant factors relate primarily to the particular work in question, the fair use analysis is not entirely outside the competence of the courts, but it is nonetheless a difficult inquiry.[150]

As we noted in the introduction, the efficiency of *ad hoc* decision making is likely to reflect the competence and information available to the court in question and this is particularly

---

[149] It might be argued that users would be reluctant to infringe copyright for moral reasons even if an infringement action was not a realistic possibility. This concern could be addressed by granting an implied licence when transaction costs were high. This licence would avoid the problem of reduced incentives to create innovate licencing mechanisms since the author would revoke the implied licence only if transaction costs could be reduced sufficiently. This solution will not work when opportunism is a threat.

[150] See Landes & Posner *supra* n.000 at 358 emphasizing the need for caution in applying the fair use doctrine.

relevant in the context of fair use. Since the fair use inquiry is a difficult one, when relevant information is not available, a court which applies an efficiency norm will have to guess at the relevant facts and sometimes it will guess wrong. For this reason I will argue that this analysis is descriptively consistent with the cases in the sense that the decisions display a concern for relevant factors, and in particular sunk costs, but I will not argue that specific decisions are always efficient. This accords with the widespread recognition that the fair use defence can be highly controversial in its application.[151]

The focus on opportunism explains why absence of a productive use weighs heavily against a finding of fair use. In general, sunk costs are unlikely when the work is purely reproductive and thus the potential for opportunism is minimal. This implies that the fact that the original author chose to bring an infringement action is in itself sufficient evidence that *ex ante* transaction costs were low. This implies, for example, that private photocopying should not be considered fair use: so long as there is no effective means of enforcing copyright in these circumstances, holding private photocopying to be an infringement will not curtail copying while still providing an incentive to develop new technologies which can lower transaction costs for licencing of private photocopies.[152]

The most prominent exception to the rule that purely reproductive use is not fair use is the decision of the Supreme Court in *Universal City Studios Inc v Sony Corp. of America*[153] in which the majority held that video-taping of broadcasts by home users for time shifting purposes was fair use.[154] Because the home users had incurred no sunk costs, there would be no scope for

---

[151] So the dissent in *Universal City Studios Inc v Sony Corp. of America* 464 U.S. 417, 475, 104 S.Ct. 774, 805 (1984) noted that 'The doctrine of fair use has been called, with some justification, 'the most troublesome in the whole law of copyright'" citing *Dellar v. Samuel Goldwyn, Inc.,* 104 F.2d 661, 662 (2nd Cir. 1939).

[152] And see *Princeton University Press v. Michigan Document Services* 99 F.3d 1381, 1996 U.S. App. LEXIS 29132 (1996, 6th Cir.) and *American Geophysical Union v Texaco Inc.* 60 F.3d 913 1994 U.S. App. LEXIS 40786 (1994 2nd Cir.), finding commercial copying of course packs and systematic copying of journal articles by a company not to be fair use. This illustrates a shift from a regime of tolerated infringement to licencing in a context where the involvement of an institutional copier makes transaction costs low enough to make licencing worthwhile.

[153] 464 U.S. 417 (1984).

[154] The majority in the Supreme Court held that time shifting was fair use because it expanded the viewing market and so benefited copyright holders who could charge more for advertising. This is a balancing argument and as such is both questionable in itself – surely the companies are better placed than the courts to determine what is in

opportunism in an action against them and the property rights approach would indicate that fair use should not apply. However, the plaintiff studios did not sue individuals who were using VCRs to tape home shows. Rather, they sued the manufacturers and distributers of the VCRs, seeking money damages and an injunction against future sale of VCRs. This raises a real threat of opportunism as the manufacturers of VCRs had no doubt incurred large sunk costs in their development and manufacture. If the plaintiffs had obtained an injunction on the sale of VCRs they would have been able to extract some of these sunk costs. As we have noted, the threat of opportunism arises because of the proprietary nature of the copyright remedy, and in holding that the defence of fair use was not available the Court of Appeals and the dissent in the Supreme Court expressed concern that the District Court and majority in the Supreme Court had improperly allowed a concern for the remedy to affect the analysis of liability.[155] And those judges who held that fair use was not available did not dismiss the problem of opportunism, but argued that it should be addressed directly in fashioning the remedy, indicating that damages or a continuing royalty, which do not give rise to the problem of opportunism, might have been appropriate relief.[156] Of course the threat of opportunism alone is not enough to make the case for a defence of fair use, since we must also ask whether *ex ante* licencing was feasible. In this case the manufacturers could have sought licences from all studios whose copyrighted works might be taped on a VCR before significant sunk costs were incurred. However, this licencing might have been difficult as a practical matter, since it would have had to have taken place before the market for VCRs was well established so that there would be considerable uncertainty in the

---

their commercial best interest – and likely to become obsolete. As Posner, Economic Analysis of Law 4th ed. (1992) points out at 42-43 changes in technology may have changed the premises on which the courts argument was based by allowing consumers to delete commercials while taping for time-shifting purposes.

[155] The dissent in the Supreme Court remarked, *supra* n.000 at 499 that "It is unfortunate that the Court has allowed its concern over a remedy to infect its analysis of liability," and the Court of Appeals 659 F.2d 963, 976 (1981) stated that "The relief question is exceedingly complex, and the difficulty in fashioning relief may well have influenced the district court's evaluation of the liability issue."

[156] Both the dissent in the Supreme Court and the Court of Appeals would have remanded the issue of remedy tot he District Court. In providing some guidance the Court of Appeals *id.* at 976 noted that "In discussing the analogous photocopying area, Nimmer suggests that when great public injury would result from an injunction, a court could award damages or a continuing royalty. This may very well be an acceptable resolution in this context." The dissent in the Supreme Court was similarly cautious: see *id.* at 499-500.

surplus which was being bargained over, and there might also have been a serious holdout problem. Thus the difficulty of assessing transaction costs makes it difficult to determine whether fair use was appropriately applied. The important point for the present purposes is that the decisions on both sides of the question show a strong concern with the problem of opportunism.

Another illustration of the concern for sunk costs is fair use protection for "incidental and fortuitous reproduction, in a broadcast or newsreel, of a work located at the scene of an event being reported."[157] If a band plays a marching tune at a public ceremony opening a new bridge which is covered by the local news and the news report incidentally captures a substantial portion of this tune, the news reporters will have a valid defence of fair use. If bargaining *ex ante* were possible, the licence fee would presumably be quite low since the tune is incidental to the work created by the news team, but once the news work is completed, the fee which could be demanded would reflect the value of the news work itself, rather than just the value of the tune to that work, as an injunction against the use of the work would prevent the news from being broadcast. If *ex ante* licencing is impossible and the news reporter protects itself by reporting the event without sound, the social loss will be greater than the value of the work protected. A powerful objection to this analysis is that the station can simply obtain a blanket licence for playing songs from one of the performing rights societies, in which case *ex ante* transaction costs will be low and the defence of fair use is not justified.[158] Nonetheless this aspect of the fair use defence shows a judicial concern for sunk costs, even though this concern was misplaced. This is arguably a case in which the courts had to guess at the relevant transaction costs, and they simply guessed wrong. It is noteworthy that the an English court has clearly held that the defence of fair use is *not* available in these circumstances.[159] This divergence of opinion may be due to different intuitions regarding *ex ante* transaction costs in the face of insufficient information.

Similarly, in some circumstances the courts may also use fair use to prevent opportunism

---

[157] Given as an example of fair use in the Registrar's Report on the General Revision of the U.S. Copyright Law (1961) reproduced in Nimmer *supra* n.000 App. 14 at 14-39.

[158] See Landes & Posner *supra* n.000 at 358.

[159] *Hawkes and Son (London) Ltd. v Paramount Film Service Ltd* [1934] 1 Ch. 593.

in the face of irrational copying by the defendants. For example, in *Mathews Conveyer Co. v. Palmer-Bee Co.*[160] the plaintiff and defendant were competitors selling various mechanical parts in competition. The defendant manufactured some parts to serve as direct substitutes for the plaintiffs (unpatented) goods and an artist preparing a catalogue for the defendant was alleged to have used a photograph of a bearing in the plaintiff's catalogue as model for a sketch of the defendant's replacement bearing, thus indirectly copying the plaintiff's engineering sketches. In holding this to be fair use, the Court stated that "[W]here the proportion taken is insignificant compared to the injury which would result from stopping a party's use of a large volume of independently acquired information, an injunction will be denied."[161] This clearly evidences a concern for preventing the plaintiff from taking advantage of the defendant's sunk costs to extract an exorbitant licencing fee. The defendant's copying was irrational in that it could have been avoided at very low cost since the artist surely knew that he or she was copying the plaintiff's work and infringement could have been avoided by *ex ante* licencing, or, more plausibly, by sketching from the defendant's bearing directly. Thus these decisions are inconsistent with the argument that fair use should only be applied when *ex ante* licencing is not feasible. However, the courts are not always purely forward looking in their decisions and these cases appear to be examples in which the opportunism was sufficiently clear that judicial second guessing of the market was perhaps not unreasonable.

A similar explanation applies to quotation of short passages in a scholarly or technical work. Scholarship is by nature cumulative, and in grappling with issues in any given field, and author must necessarily refer to earlier work. In many areas of scholarship, verbatim quoting of some portions of a previous work is necessary. Licencing would be possible before the second

---

[160] 135 F.2d 73; 57 U.S.P.Q. (BNA) 219 (6th Cir 1943)

[161] *Mathews Conveyer id.* at 84-85, citing *Dun v. Lumbermen's Credit Ass'n* 144 F. 83. (7th Cir. 1906) which similarly illustrates the importance of sunk costs and opportunism. In *Dun* the defendant prepared credit reports of lumbermen while the plaintiff prepared credit rating for business men more generally. Within its narrow compass, the defendant covered more individuals and provided much more information than did the plaintiff. It was the practice of the defendant to collect its own information and only to cross check it against Dun's in some instances, but on some few occasions it appears that employees of the defendant may have used Dun's information improperly. The Court held that even if this was the case, the use was fair: "[T]he proportion [improperly used] is so insignificant compared with the injury from stopping appellees' use of their enormous volume of independently acquired information, that an injunction would be unconscionable" (at 84-85).

work is started, but in scholarship the sunk costs have already been incurred by the scholar in becoming familiar with the field, that is, before deciding to begin any particular work.

Another factor to be taken into account in determining whether the use was fair is the nature of the copyrighted work. In part this means that unpublished works are deserving of greater protection than published works. This is relevant to the assessment of the net benefit of dissemination as it shows a concern for the difficulty of quantifying the value of privacy. While the consideration of the threat of opportunism adds nothing to the analysis of this factor, it is not inconsistent with it. However, consideration of the nature of the work is also sometimes taken to mean that the scope of fair use should be greater for informational works.[162] As should be evident from the discussion of copyright in facts, this is inconsistent with the property rights analysis.

A second possible inconsistency with the property rights approach is the rule that copying of literal software code for the purpose of reverse engineering ideas may be considered fair use so long as only the ideas are used in the defendant's work.[163] This is defended on the basis of a balancing argument: if the idea/expression dichotomy accurately strikes the balance between incentives to create and restrictions on dissemination, then a creator should not be permitted to alter this balance. But if exploring the creator's idea requires easily provable copying of expression, the author has in effect succeeded in creating a clearly definable property interest in the idea. The property rights analysis therefore indicates that protection should be granted. Put another way, this is a case in which the line between idea and expression should shift as technology has allowed clearer definition of property rights than was previously possible.[164]

---

[162] NIMMER ON COPYRIGHT § 13.05[A][2][a]

[163] See *Sega Enterprises Ltd. v Accolade Inc.* 977 F.2d 1510 (9th Cir. 1992) at 1526-27 and E.U. Directive on the Legal Protection of Computer Programs (91/250/EEC) Art. 5. For a critique and discussion of cases to the contrary see Arthur R. Miller, Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?, 106 Harv. L. Rev. 977 (1993).

[164] There is a distinct argument against permitting an author to use copyright to protect ideas which is arguably consistent with property right theory. If copyright protection can be used in this fashion to protect ideas, the effective duration of protection is the extended copyright period rather than the more restricted patent period. If the patent term is more appropriate for computer program related ideas (which is by no means clearly established) then denying copyright protection may be justifiable as a practical second best even if the copyright form of protection is desirable. See discussion of the relationship between the term of protection and the form *supra* n.000.

III.C Derivative Works

Unauthorized creation of derivative works is of course an infringement of the copyright in the original. There is a striking contrast between the balancing and property rights approaches to copyright in explaining this basic point.

The balancing approach must explain why the increased incentives to create original works outweighs the decreased incentive to produce derivative works, so as to produce a net benefit if the market for derivative works is protected. Landes and Posner's treatment of the issue illustrates the informational difficulties associated with the balancing approach.

First consider the arguments relating to incentives for creation. Landes and Posner argue that contrary to the obvious suggestion, the justification for protection of the market for derivative works is *not* that this is required to enable the original author to recoup his fixed costs, since the derivative works are not demand substitutes and do not reduce the author's revenue in the primary market.[165] They concede that the inability of the author to capture revenue from derivative works will be taken into account in the decision to create the original work, but imply the effect on incentives for creation will be small.[166] This is difficult to credit. We know that in the movie industry, for example, licencing of various spin-off products and tie-in promotions generate revenues running into the tens and even hundreds of millions of dollars, which may amount to many times the cost of production of the movie in question. There can be no real doubt that studios, which generally have senior executives whose main responsibility is marketing derivative works, take such earnings into account in projecting revenues.[167] While the

---

[165] *Supra* n.000 at 354.

[166] They argue, *supra* n.000 at 354 that "the conferral of the right is not necessary in order to prevent [the author's] sales from being driven to zero. And since it is not certain that *any* copyright protection is necessary to enable publishers and authors to recover their fixed costs, it would be speculative to conclude that without control over derivative works authors and publishers would not be able to cover the fixed costs of the original work..." This seems to concede the speculative nature of the balancing approach to copyright protection generally.

[167] See e.g. Behemoths Line up to License on 'Park' Row. Brandweek Magazine, 18 Jan 1993, p.1 (reporting that "Jurassic Park" is predicted to bring in over $100 million in licencing revenues); 'Lion' is new king of licensing jungle, Advertising Age, July 4, 1994 ("The Lion King" expected to top $1 billion in merchandise sales); Licensemania, Direct Marketing Magazine, April 1994, p.26 ("Licensed products resumed their upward climb in 1993 with a 7 percent increase over 1992's figures."); "Stars Wars" empire shows new strength, Advertising Age, December 5 1994 ($500 million in licence revenues); Hollywood's star rising for marketers, Advertising Age, August 8, 1994, p.3 (general importance of tie-in marketing); Brand builders, Brandweek, 27 March 1995, p.20

bulk of the licencing fees are generated by a few blockbusters, these fees will nonetheless be expected to have a significant effect on overall incentives to make movies by their influence on expected returns. In the movie industry this is amply confirmed by anecdotal evidence which indicates that a few hits support the production of a large number of money losers.[168] This casual empiricism indicates that Landes and Posner are almost certainly wrong on this central empirical point in respect of the movie industry.

But we cannot conclude from this that the argument that the need to provide incentives to create original works justifies the protection of derivative works. We know that it provides significant incentives in the movie industry, but without more empirical evidence, it is difficult to say, on the basis of intuition alone, whether this is true in other sectors. In the first place, we haven't even addressed the issue of the cost of protecting that market in the movie industry, which is necessary to decide if there is a net benefit. Even if we can conclude that protection of the market for derivative works creates a net benefit in the movie industry, it may be that it creates net losses in other sectors, and that these losses outweigh the gains in the movie industry. The point, then, is not that Landes and Posner are wrong in arguing that the need for incentives is not the primary justification for protection of the market for derivative works. It is that the issue of whether such incentives are necessary is an empirical issue where our intuitions are unreliable at best.

Having dismissed the straightforward incentive argument, Landes and Posner offer two main reasons for protecting the market for derivative works. They begin by they noting that derivative works themselves should be protected, offering the standard argument that free-riding would otherwise inhibit the production of derivative works. They then argue that granting copyright in the derivative work to the original author, rather than the author of the derivative work, will reduce transaction costs. By way of illustration they suggest that if the copyright in

---

(role of studio marketing executives in promoting tie-ins); nor are such deals restricted to major studios: see "Spike strikes again" Forbes, April 11 1994 (pre-release licencing for Spike Lee's *Crooklyn*). Nor is the success of derivative works is a purely modern phenomenon: Libott *supra* note 000 states that "throughout the Nineteenth Century clever dramatists (or, more likely, their producers) made substantially greater profits out of the successful novels of the period than the writers of the original books" (at 744).

[168] See e.g. "You're not in Kansas any more," The Economist, Feb. 4th, 1995, p.57.

the Russian original and the English translation of *The Brothers Karamazov* were held by two different copyright holders, namely the heirs of Dostoevsky and the American translator respectively, a publisher who wanted to bring out a new edition of the translation would have to deal with two copyright holders. Transaction costs, they suggest, are lowered if one person holds both copyrights.[169] The difficulty with this argument is that if one person holds both copyrights, transaction costs are raised in the first instance as the translator must negotiate with the holder of the copyright in the original version, since the translator will clearly not undertake to translate the work without some assurance of compensation. Just as many transactions occur under either rule, and the efficient pattern will depend on the relative costs of the various transactions, and on the proportion of translations which have second editions produced. If we assume that not all translated works are brought out in new editions, and that all the relevant transaction costs are roughly equal, then transaction costs are minimized if the translator holds the copyright in the translation. In any event, on their assumption that revenues from licensing fees for the derivative works are not needed to supply incentives to create the work in the first place, transaction costs are clearly minimized if the author of the derivative work and not the author of the original, since this eliminates the costs of negotiation between the author of the derivative work and the author of the original work.

An alternative argument made by Landes and Posner is that if only the translator held the copyright, the original author would inefficiently delay publication of the original work until she had also prepared the translation. This point is at odds with their argument dismissing original author incentives as a reason for original author copyright in derivative works, as it clearly requires that the original author take into account the revenues from the derivative works in creating the first work. Further, the benefit of reduced costs to subsequent works was a primary argument they made against allowing copyright in ideas, and it also appears relevant here: since it is unlikely that the original author would generally have sufficient resources to create all possible derivative works, the distortion from any delay would be offset by the fact that after release of the original and whatever derivative works occurred to and were within the organizational means of the author/publisher, creators of subsequent derivative works would be

[169] *Id.*. at 355.

able to use the original work at marginal cost, thereby reducing the cost of those subsequent derivative works.

Landes and Posner's analysis of the treatment of derivative works illustrates the complexity and contingency of the balancing analysis. Their argument depends crucially on strong assumptions regarding empirical issues, namely the relative size of various transaction costs and incentive effects. Landes & Posner accept that the balancing argument requires guessing about the magnitude of incentive effects, but they argue that in general "the courts have made intelligent guesses."[170] But Landes & Posner's own guesses range from the debtatable, to the almost certainly wrong. Why should we suppose that the courts are doing significantly better?

The property rights argument, in contrast, is trivial. A derivative work is, by its nature, a work which copies the expression of an earlier work in the creation of a new work. Since there is no controversy over the fact of copying, property rights are well defined, and the copying should be considered an infringement.

Any theory will be stretched in border-line cases, like the idea/expression dichotomy. But a plausible theory should provide simple explanations for clear-cut cases. The property rights approach passes this test the context of derivative rights, whereas the balancing approach, at least as applied by Landes and Posner, does not.

IV CONCLUSION

Intellectual property is property. It can be owned, bought, sold, given away – in short, it shares many characteristics with ordinary tangible property. But it is also "intellectual". It is a property right in an intangible. It is a "public good" which can shared by as many who want it, without diminishment. This distinguishes it sharply from tangible property – intellectual property is in some sense a true horn of plenty. It is no surprise that analysis of copyright has tended to focus on what is unique – the 'intellectual' or 'public good' aspect – in preference to

[170] *Supra* n.000 at 336.

the mundane property aspect.[171]

In this Article I have argued that the this emphasis, while understandable, is misplaced. Intellectual property law is not really about optimizing the law to take best advantage of intellectual property's public goods characteristics. It best understood as being about the more modest task of ensuring that property rights are well defined. The registry systems and priority rules of the law of personal and real property are concerned largely with providing efficient notice to the world of the ownership interests. I have argued that the key doctrine of the law of intellectual property serve exactly the same purpose: to ensure that the public dealing with intellectual property has clear notice of the ownership interests. The law of real and personal property clearly does not seek to directly mandate an optimal allocation of resources, and most economists, certainly, would recoil from the suggestion that it should. I have argued that we should be equally sceptical of the notion that the law of intellectual property seeks to achieve an optimal balance in the initial allocation of property rights is a popular one. In both cases, the aim of the law is and should be to facilitate private trade by providing clearly defined property rights.

✦ ✦ ✦

[171] So, Landes & Posner *supra* n.000at 326 remark that "A distinguishing aspect of intellectual property is its 'public good' aspect."