The Economics of Intellectual Property

Centre for Economic Analysis of Property Rights

Follow this and additional works at: https://ir.lib.uwo.ca/economicsceapr_el_wp

Part of the Economics Commons

Citation of this paper:
THE ECONOMICS OF INTELLECTUAL PROPERTY
March 31 - April 2, 1983
Preliminary Drafts

CENTRE FOR ECONOMIC ANALYSIS OF PROPERTY RIGHTS

Economics and Law Workshop Papers
83-08A

University of Western Ontario

THE UNIVERSITY OF WESTERN ONTARIO
ECONOMIC ANALYSIS OF INTELLECTUAL PROPERTY RIGHTS

CONFERENCE SCHEDULE

Thursday, March 31, 1983

6:00 p.m. Buffet Dinner for Early Arrivals

7:15 p.m. Registration and Coffee

8:00 p.m. Session I

A. "Property Rights and Invention" by Steven N. S. Cheung
   (University of Washington and University of
   Hong Kong)
   Discussant: George Priest (Yale Law School)

10:00 p.m. Cash Bar

Friday, April 1, 1983

8:00 a.m. Breakfast for Participants Staying at Spencer Hall

9:00 a.m. Session II

A. "The Patent Term, the Scope of the Patent Right, and the Definition of Competition" by Donald
   G. McFetridge and M. Rafiquzzaman (Carleton
   University)
   Discussant: Roger Beck (University of Alberta)

B. "Patents and Market Power" by Edmund W. Kitch
   (Virginia Law School)
   Discussant: F. M. Scherer (Swarthmore College)
Friday, April 1, 1983 (cont'd.)

10:30 a.m. Coffee Break

11:00 a.m. Session III
   A. "Patents, Licensing, and Antitrust" by Christopher D. Hall (Simon Fraser University)
      Discussant: Louise Séguin-Dulude (École des Hautes Études Commerciales)
   B. "Copyright Collectives" by Douglas Smith (Carleton University)
      Discussant: A. Hollander (University of Montreal)

12:30 p.m. Lunch

2:00 p.m. Session IV
   A. "Home Taping and Copyright Payments" by James Keon (Consumer and Corporate Affairs)
      Discussant: Robert E. Dansby (Bell Labs)
   B. "Copyright and Reprography" by Stan Liebowitz (Rochester Graduate School of Management Studies)
      Discussant: Bill Bishop (Wolfson College, Oxford)

3:30 p.m. Coffee Break

4:00 p.m. Session V
   A. "Copyright and Computer Software" by John Palmer (University of Western Ontario, Centre for Economic Analysis of Property Rights)
      Discussant: Stephen Margolis (North Carolina State University)
   B. "Personality and Personality: The Right of Publicity" by Howard P. Knopf (Attorney; Toronto)
      Discussant: Barry Torno (Attorney; Cassels, Brock; Toronto)

6:00 p.m. Dinner
Saturday, April 2, 1983

8:30 a.m.  Breakfast

9:30 a.m.  Session VI

A. "The Origins of Canada's New Stance on Intellectual Property" by David Bond (British Columbia Senior Advisor)

B. "Canada's Present Stance on International Obligations for Intellectual Property" by Fenton Hay (Consumer and Corporate Affairs, Bureau of Policy Analysis)

C. "International Strategic Considerations in Copyright Law Design" by John Whalley (University of Western Ontario)

Discussant: Stan Liebowitz (Rochester Graduate School of Management)

12:00 p.m.  Lunch
PROPERTY RIGHTS AND INVENTION

Steven N. S. Cheung*
University of Hong Kong

Ideas are highly valuable resources. Yet it was not until 1471, in Italy, that a system of patent rights was devised to protect intellectual properties. Patentable inventions constitute only a small subset of intellectual properties. In terms of complexity, however, the patent system today far surpasses any other method of protection of invention (such as trade-secret arrangements or copyright). This paper critically reviews early and existing economic arguments relative to the patent system. The intent is not only to identify logical and conceptual errors where they occur, but also to marshal those aspects of property rights in ideas with which economists have largely dealt.

* For a subject as controversial as this, it is difficult to take advantage of the many and varied comments. However, I am grateful to the following for suggestions and queries which have helped greatly in clarifying my thinking: Kenneth Arrow, Yoram Barzel, Gary Becker, Ward S. Bowman, Jr., R. H. Coase, Christopher D. Hall, Jack Hirshleifer, Carole Kittl, Ilona Melstrads, Roif Piekars, Dean Worcester, and Benjamin Yu.

The paper has also benefited by its presentation at UCLA, the University of Chicago, and Western Economic Association. For assistance in its preparation, thanks are due to Lina Tong, Charlotte Twilight, Michael Sharp, and Marion Impola.

Financial support throughout has been provided by the National Science Foundation Grant Number PRA76-19804.
Interestingly enough, on the subject of establishing private property rights over resources, the curiosity of economists has been more piqued by those arising from invention than from most other activities. This may be attributable to that the delineation and enforcement of rights to an entity as elusive as an "idea" capture the imagination. In the economics of property rights, no more disparate set of arguments can be found than those dealing with the functions or implications of a patent grant.

With perspective sharpened by distance, we can now detect four roughly separable groups of arguments. One view — advanced by Bentham (1795) and shared by Say (1803), Mill (1848), and Clark (1907) — holds that patent rights are absolutely necessary to encourage inventions. A second view, advanced by Taussig (1915) and shared by Pigou (1920), maintains that a system of patent rights is largely superfluous.\(^1\) Third, Plant, with modern followers, argues that a patent system is detrimental. Finally, Arrow (1962), utilizing in part the works of Hotelling (1938) and Samuelson (1954), argues that although property rights in ideas are clearly useful, they are nonetheless inferior to direct government investment in inventive activities.

---

\(^1\) The dates in parentheses correspond to the year in which the works were first published, and are not necessarily the editions to which page references are made later in this paper.
I. The Early Views: From Bentham to Pigou

With reference to inventions, Jeremy Bentham noted that "an exclusive privilege is absolutely necessary in order that what is sown may be reaped." Thus, an inventor "who has no hope that he shall reap will not take the trouble to sow." This, together with Bentham's assertion that the patent system "costs nothing," accounts for the early thesis that instituting property rights over ideas enables society to gain something for nothing. J. B. Say supported the argument by noting that a patent interferes with nothing to exist, while J. S. Mill asserted that it added only a trivial cost. By the time the argument reached J. B. Clark, the something-for-nothing thesis had become almost firm: "If the patented article is something which

---


3 Ibid.

4 Ibid.

5 "Privileges of this kind no one can reasonably object to; for they neither interfere with, nor cramp any branch of industry, previously in operation. Moreover, the expense incurred is purely voluntary; and those who choose to incur it, are not obligated to renounce the satisfaction of any previous wants." J. B. Say, A Treatise on Political Economy [1834] (New York: Augustus M. Kelley, 1964), p. 182.

6 "[The patent grant] is not making the commodity dear for his benefit, but merely postponing a part of the increased cheapness, which the public owe to the inventor, in order to compensate and reward him for the service. That he ought to be both compensated and rewarded for it, will not be denied." J. S. Mill, Principles of Political Economy, 2 volumes [5th London edition, 1862] (New York: D. Appleton and Co., 1864), 2:548.
society without a patent system would not have secured at all -- the inventor's monopoly hurts nobody ... his gains consist in something which no one loses, even while he enjoys them.  

Yet even early economic analysis of patents began to uncover defects in the "something-for-nothing" theory. Clark himself did not wholeheartedly support the thesis that the patent system had no side effects on the economic system at large, and the something-for-nothing view later was attacked by Arnold Plant, whose ideas will be discussed later. But one major flaw in the thesis which has passed largely unnoticed is the failure of the early writers to examine certain unavoidable complexities of any patent system -- specifically, to consider the appropriate subject and scope of legitimate patent grants. That is to say, the criteria of patentability were ignored.

Even if the delineating and enforcing of rights to an invention were to involve no costs, the troublesome question of what ideas should be granted patent protection must be faced. In one extreme, there is

---


8 See the section entitled, "How Patents May Cause an Increased Displacement of Laborers," ibid., pp. 265-6.
nothing new under the sun. In another extreme, every different combination of ideas or every different application of an idea constitutes a new idea. In specifying the criteria of patentability, the designers of any patent system must select a position somewhere on the spectrum marked by these extremes. In the United States, for example, the "discovery" of principles of nature is not patentable, a rule to which Bentham would have objected. On the other hand, a different use of an unpatented product is patentable, as for example, the use of unpatented arsonic acid to control crabgrass. Yet the early writers paid no attention to the question of the appropriate criteria of patentability.

Even if there were no difficulty in determining what is new and patentable, to what extent must patent claims be delimited? Again, the early writers gave us no answer. Consider a cost-saving device. One inventor comes up with a device which will save $5 per unit of output whereas another saves $3 per unit. Are we to grant protection to both, to the superior device only, or to the superior device with

---


10 For this specific case "the court held that it was proper for the patentee-licensor to sell the unpatented product and charge a royalty for its use under the patent, with the burden of proving nonuse on the purchaser." Thomas E. Costner and Harold Einhorn, Patent Licensing Transactions, Patent Law and Practice, vol. 1, rev. (New York: Matthew Bender, 1974), pp. 1-14. The case was Noll v. O. M. Scott, 169 U.S.P.Q. 336 (S. D. Ohio 1971).
the restriction that the patentee cannot extract a royalty in excess of $2 per unit? Or consider a new product consisting of parts A, B, and C. Are we to issue patents to AB, AC, and BC? Regardless of the ultimate verdict on the merits of patent protection, the early writers were premature in lauding its efficiency without considering these issues.

Contrary to the something-for-nothing view, F. W. Taussig argued that the patent system is roughly nothing-for-nothing. Although he placed a high social value on the advancement of ideas, he regarded invention as arising mainly from a philosophical "instinct of contrivance":

One thing stands out conspicuously: the race of contrivers and inventors does obey an inborn and irresistible impulse. Schemes and experiments begin in childhood, and persist so long as life and strength hold. It matters not whether a fortune is made or pecuniary distress is chronic.\(^{11}\)

Taussig accordingly concluded that "the patent system ... is a huge mistake."\(^{12}\) He modified this view to the extent of adding that although patent legislation will not significantly increase inventive activity, it may direct the activity "into channels of general usefulness."\(^{13}\)

A. C. Pigou agreed with Taussig, but he added a new dimension to the discussion by considering the aspect of social cost:

We may set out a number of instances in which marginal trade net product falls short of marginal social net product, because incidental services are performed to third parties from whom it is technically difficult to exact payment ... [M]ost important of all [is the case where] resources and activities [are] devoted alike to the fundamental problems of scientific research, out of which in unexpected ways discoveries of high practical utility often grow, and also to


\(^{12}\) Ibid., p. 18.

\(^{13}\) Ibid., p. 51.
the perfecting of inventions and improvements in industrial processes. These latter are often of such a nature that they can neither be patented nor kept secret, and therefore, the whole of the extra reward which they at first bring to their inventor is very quickly transferred from him to the general public in the form of reduced prices. The patent laws aim, in effect, at bringing marginal trade net product and marginal social net product more closely together. By offering the prospect of reward for certain types of invention they do not, indeed, appreciably stimulate inventive activity, which is, for the most part, spontaneous, but they do direct it into channels of general usefulness.\textsuperscript{14}

For a case which he regarded as "most important," it is puzzling that Pigou attempted no demonstration at all of how, under his presumption of "spontaneous" invention that cannot be kept secret, private and social costs would diverge without patent laws. Such an attempt at explanation would have failed, for if inventions are truly "spontaneous" and if knowledge of them cannot be kept secret, a divergence between private and social costs results from the presence of patent laws, not from their absence. Let us examine Pigou's reasoning more closely to see why this is so.

Pigou's approval of the patent system, despite his belief that it does not appreciably stimulate invention, rested on the premise that the patent system would increase society's welfare by directing invention into "channels of general usefulness." Of two possible interpretations of that phrase, each presents difficulties. One argument is that a patent system would encourage the application of an invention to production. Yet once an invention is spontaneously produced, any royalty payment made possible by patent protection must at some margin inhibit its use. Even

if the inventor chooses to use the idea in production himself, he and his competitors combined would have used the invention still more had no patent been granted. Thus, under Pigou's specific assumptions, patent laws would contract rather than expand social "welfare." A second interpretation is that the patent system will direct the invention to more valuable uses. But the distinctive characteristic of an already developed idea is that it is amenable to concurrent uses by multiple users in multiple ways at no additional cost. In short, the marginal cost of accommodating an additional use is zero. If the invention is spontaneously (costlessly) produced, the most efficient allocation again is to grant free access to all. In either case, the patent system would create a divergence between private and social costs.

Thus, under his own assumption of spontaneity, Pigou was wrong in arguing that patent laws would increase society's welfare. If the invention is to emerge regardless of pay, then by its nature of being amenable to concurrent uses any protection which inhibits its use will, by Pigou's own measure, generate economic waste. In the strictest sense, a "spontaneous invention" would be one produced without cost, in which case decision no longer involves choice amidst scarcity, and inventive behavior lies outside the domain of economic analysis. But can it really be true that inventions are "spontaneously" produced?

Consider what Taussig and Pigou would agree is the purest case of a spontaneous invention: discovery by accident. Could an economist accidentably formulate the theory of relativity, or could Einstein, without a change of profession, come up with the Coase theorem? Scientists who engage in serious research may hope for some lucky
accidents, but to increase the chances of "accidental" discoveries they must specialize. That very specialization entails costs which necessitate economic choices. The high degree of specialization of various research-and-development laboratories in the United States,\textsuperscript{15} the emergence of development contracts even before the creation of the patent system,\textsuperscript{16} and the prevalence of employed inventors,\textsuperscript{17} all attest the improbability of purely "accidental" invention. Indeed, Pigou himself was inconsistent when, in another work, he discussed at length how inventive activity will vary in different settings.\textsuperscript{18} How can an invention be spontaneous if it is also a product of economic inducement?

In refuting Pigou's assumption of spontaneity, however, we reopen the question of whether patent laws are desirable in a world of deliberate, volitional invention — a world in which inventive activity is largely determined by choice in the presence of scarcity.


\textsuperscript{17} See Fredrik Neumeyer, \textit{The Employed Inventor in the U.S.: R \& D Policies, Law and Practice} (Cambridge: M.I.T. Press, 1971). It is interesting to note that, as a percentage, the number of employed inventors has risen sharply in the United States since 1901. In the period of 1956-60, patents issued to employed inventors constituted 63.6 percent, as against 36.4 percent for outside inventors. See Jacob Schmookler, \textit{Inventions and Economic Growth} (Cambridge: Harvard University Press, 1966), p. 26.

II. The Prevailing Views: Plant and Arrow

After the debate once opened, it would seem only a matter of time until someone would describe the patent system as promoting something-for-something — that is, achieving certain ends only at certain costs. As the principal architect of that thesis, Arnold Plant criticized Say and Clark for failing to consider that patents draw scarce resources into the production of patentable ideas.\(^{19}\) While acknowledging that advocates of patent laws hope to simulate invention by increasing the economic reward capturable by inventors, Plant concluded that patent laws, on balance, are detrimental. He argued that with or without a patent system the price incentives in an unfettered market lead to the most efficient economic result in the production of ideas/inventions just as in the production of other (tangible) commodities. While recognizing some spontaneous invention, he insisted that its output is not very important in volume or in kind.\(^{20}\) Feeling that price movements and business conditions (reflecting consumers' preferences) adequately stimulate invention, Plant viewed patent laws as creating artificial monopoly rewards for inventors of patentable ideas. Thus their net result, in his view, was to reduce society's economic welfare through creating monopolies. To appreciate Plant's reasoning, consider the wheel, an unpatented invention. It requires little stretch of the imagination to envision the damage to society if the utilization of the wheel (or, more properly, the axle) had been monopolistically restricted under the protection of an enforceable and perpetual patent grant.


\(^{20}\) Ibid., p. 42.
Plant argued further that the monopoly reward implied by the patent grant would induce inefficient invention, diverting scarce economic resources into production of inventions deemed "patentable" under necessarily arbitrary rules. In short, he envisioned the lure of patent monopolies as leading to misallocation of resources, distorting consumers' preferences in a free market.

Monopoly pricing was but one of several defects Plant alleged in a patent system. He noted that "lotteries in open competition there may well be; but the lottery of the patent system awards but one prize, and that a monopoly, while those who subscribe most of its value may be precluded from qualifying for the prize."21 As a result, "the grant of monopoly renders almost nugatory the labours of all the rest."22 This turns out to be the fountainhead of the modern thesis of "rush to invent" and of "wasteful duplication of invention." The fallacy of this thesis will be identified in another paper.23

Central to Plant's argument regarding monopoly rewards attendant upon patent grants was his differentiation of property rights in patents from those in other resources:

It is a peculiarity of property rights in patents (and copyrights) that they do not arise out of the scarcity of the objects which become appropriated. They are not a consequence of scarcity. They are the deliberate creation of statute law; and, whereas in general the institution of private property makes for the preservation of scarce goods, tending ... to lead us to "to make the most of them," property rights in patents and copyright make possible the creation of a scarcity of the products appropriated which could not otherwise be maintained.24

21 Ibid., p. 46
22 Ibid.
23 See Steven N. S. Cheung, "The Right to Invent and the Right to an Invention," unpublished manuscript.
In only one situation did Plant see the value of a patent system: as an incentive to create inventions requiring costly and prolonged research which the market would not reward in the short run. Nevertheless, asserting that this class of research represents "exceptional circumstances," he contended that "a patent system applicable to inventions in general clearly cannot be justified ... by exceptional circumstances of this kind." Plant further argued against the patent system that a grant tends to inhibit fruitful subsequent work because it blocks improvements based on the original patent.

The phrase "a patent system applicable to inventions in general" (a condition central to Plant's argument) is decidedly ambiguous. A broad interpretation could extend protection to the laws of Newton and Mendel. This is not what Plant meant, for he also used the term "patentable inventions" which implies that the patent system is necessarily discriminatory in nature. To eliminate contradiction, one must then select criteria of patentability which are to be uniformly applied to "inventions in general." Roughly speaking, this is the nature of various patent systems today. But the criteria of patentability are subject to alterations over time and place; thus, ignoring transaction costs, one could construct a set of criteria to render Plant's argument wrong.

Consider the case of the wheel, for which certain designs are patentable. Now any child can draw a circle, and a round object can be put to uses so trivial that however "new" they may be, they hardly meet the criterion of novelty. On the other hand, a patent system may

---

25 Ibid., p. 43.
encourage both the usefulness and the speed of development of certain modifications and applications related to the wheel such as the axle, gear, and ratchet. In the virtually endless array of things which may be developed from a round object, each — at some margin — will be similarly responsive to development under the protection of the patent system if the claims are sufficiently narrowly defined.

Thus it is empty to argue the economic value of a patent system without specifying the criteria of patentability involved: to evaluate these criteria and their enforcement is to evaluate the patent system. This task can be performed, but its difficulty and tedium are enormous. This is especially true when the costs of transaction (and of enforcement) are significant, but these constraints were not recognized until Kenneth Arrow tackled the subject.

Return to Plant's thesis that a patent system, in granting monopoly rights, creates scarcity. To analyze this view, it is helpful to distinguish between inventions that would have been produced in the absence of any property-rights protection and those that would not have been produced without protection. Plant's view may be true with respect only to the former class. For such inventions, the patent grant will, under certain pricing arrangements, inhibit the widespread use of an already developed idea. "Scarcity" may then be said to have been "created" by protection.

However, for ideas which would not have been produced without some form of protection, Plant's view is in error. Suppose gold may be discovered through difficult and costly search in the mountains; in the meantime the hidden gold is of no use to society. Can we say that property rights

26 Valiant efforts in that direction are represented by the authors and works cited in fn. 9.
granted to its finder create scarcity? Or consider a piece of wasteland that has no market value. If someone by assiduous tilling converts that acreage into valuable property, can we say that the assignment of property rights to the land is creating scarcity? As with any commodity which would not have been produced at all in the absence of property rights, the scarcity is neither of gold, nor of wasteland, nor of potential ideas; rather, in each case the scarcity lies in the resources required to bring these prospects to useful reality.

Protection of rights to an idea, whether by patent, by trade secret or by any other device, will give rise to the same monopoly problem visualized by Plant. Even a college professor who succeeds in protecting his research, not by copyright but simply by establishing a reputation, is in effect a monopolist. This leaves open the question of what type of invention will take place without any kind of protection.

First, even if there were such a thing as "spontaneous invention," it is regarded by Plant as negligible. A second type of invention devoid of protection is that which results merely from the lively curiosity of the researcher, but geniuses like Mendel are in short supply. A third type generates sufficient reward for the inventor in his own use; most householders and amateur gardeners tend to make innovative use of their toolboxes and implements. The resources which an inventor of this type is willing to commit will not exceed the expected return to himself; however, although he gives no thought to possible social benefits from his work, imitation by friends and neighbors will result in marginal social gain greater than marginal social cost. "Underinvention" is implicit in this situation.
Let me classify as *intramarginal* those inventions which would have emerged without protection and as *marginal* those that require the inducement of protection for their discovery. Assume, as Plant did, that all patents enable patentees to obtain monopoly rents (although this is not necessarily true). Under these assumptions, protection by either patent or trade secret allows society to gain from the marginal inventions since, lacking protection, they would not have been produced at all. By granting protection to the intramarginal inventions which would have been created in any case, the patent or any other protection system causes no waste so long as perfect price discrimination or some form of lump-sum extraction governs their use; rent is simply redistributed from consumers to inventors. Whether the protection system is efficient, therefore, depends on the pricing behavior of the patentees and (if an inefficient pricing practice is adopted) on whether the resultant economic waste more than outweighs the gain from the marginal patents.

Interestingly enough, the issue at stake in Plant's arguments is precisely the same as that in the something-for-nothing thesis advanced by Bentham and supported by Say, Mill, and Clark, although Plant concludes that the patent system is harmful whereas Bentham considers it indispensable. The crossroads-point is the scope of patentability or the scope of establishing rights to ideas. If a protection system can be designed to grant varying degrees of protection to different types of invention, then to say that any patent protection is harmful must be wrong. On the other hand, unless the costs of transaction are low enough to permit varying lump-sum extraction to encourage widespread
use of ideas, the scope of idea-protection may be extended to a point where the constriction of use creates social loss.

With or without a protection system, an inventor (or an entrepreneur underwriting inventive activities) will base his calculation of the rate of return on his estimation both of potential success and of his ability to capture the returns. What a patent system does is to raise the general level of his expected return. This tends to increase research effort until at the margin the expected rates of return equal those in the highest-valued alternative employment of resources. Inventions are then produced under a different cost-reward system, promoting some valuable inventions which would not otherwise have become available. This changed cost-reward system produces a series of unresolved questions: whether the system is founded on a well-defined scope of protection (or whether such a clear definition is even possible); what relationships exist among various protection systems such as patents, trade secrets, trademarks, and copyrights; what role is played by the costs of transaction in each system, separately and jointly; and how well the systems, in practice, conform to economic criteria, together with the implications of that finding on resource allocation.

Faced with that myriad of unanswered questions about property rights and inventions, some economists have despaired of the patent system and suggested direct government investment in invention as closer to the "optimal" solution. In particular, Kenneth Arrow and others have claimed that underinvestment in invention is inescapable with or without a patent system and that this problem can best be mitigated by expanded government investment in innovative activities.
In an important work, Arrow acknowledges that the patent system will encourage invention. However, he notes three main reasons for failure to achieve "optimality" in these activities: uncertainty, indivisibility, and inappropriability.\textsuperscript{27} In the case of uncertainty, he asserts that market failure stems primarily from non-neutral risk preferences of inventors and identifies risk aversion as one major cause of underinvestment in invention.

While he sees insurance and other mechanisms (e.g., securities such as common stocks and money) as means of separating risk-bearing and productive functions and thus approaching closer to the efficient allocation of inventive resources, he also points out the limitations of these insurance or risk-spreading devices. In particular, he designates a "moral factor" as an inherent limit, whether on fire insurance or on "invention insurance": to the extent, he says, that the insured is insulated from risk, he has less incentive to avoid the associated failure.\textsuperscript{28}

On this issue, we may interpret Arrow's moral factor as simply a problem of transaction costs. If one could costlessly enforce the insurance contract -- discovering at zero cost any diminished vigilance or effort by the insured -- the problem would not exist. It is the cost of enforcing such contracts that gives rise to the moral hazard.


\textsuperscript{28} Ibid., pp. 612-14.
Indivisibility and inappropriability, the other two causes of underinvestment in invention discussed by Arrow, are different sources of a single problem, that of capturability of returns from invention. Their relation to property rights in inventions is clear and can be generalized: inventions pose problems in appropriating or capturing returns even under patent protection because of the presence of transaction costs. Therefore in Arrow's view underinvestment will occur, providing a strong case for government finance of research.

Consider first the problem of indivisibility, which inhibits marginal-cost pricing. In the present context, "indivisibility" refers to the inability of marginal-cost pricing to divide the full production cost among the users of a good (or an idea). Whenever average costs are decreasing, pricing by marginal cost presents a problem. This seems particularly true for the use of an idea, when the marginal cost of accommodating an extra user is virtually zero. Arrow argues that from the welfare point of view, an invention should be made available to users free of charge. But a zero price for the use of an invention removes any incentive for investment in research. The resultant paradox cannot be more clearly expressed than in Arrow's own words:

> In the free enterprise economy, inventive activity is supported by using the invention to create property rights; precisely to the extent that it is successful, there is the underutilization of the information ... [The] problem, then, is that in the free enterprise economy the profitability of invention requires a non-optimal allocation of resources.  

---


This line of argument was attacked earlier by R. H. Coase in his discussion of the marginal-cost controversy and later by Harold Demsetz in a lengthy critique of Arrow's work.

Leaving these criticisms aside, the problem Arrow visualized seems to stem from the fact that an idea can be simultaneously employed in different uses by multiple users. Thus the marginal social opportunity cost of accommodating an extra user or an extra use is zero, and if the invention is produced without cost, then maximum social gain will indeed be attained if the invention is made freely available to all users. That this problem is one of capturability of returns becomes evident when one notes that if the costs of transaction were zero, then some form of perfect price discrimination or lump-sum extraction would insure that the marginal price be zero. Pricing at zero marginal cost for the use of an invention does not preclude economic reward to the inventor. The capture of intramarginal rent or consumer surplus through lump-sum charges would then compensate for the cost of invention, and competition among inventors would drive the capturable rent to the point that the expected rate of return would equal that in alternative use of resources.


Such intramarginal pricing has been regarded by economists as largely wishful thinking. The significance of transaction costs in the marketing or licensing of ideas can no more be denied than can the prevalence of royalty rates per unit of output in most patent licenses. However, the great complexities of the rate structures and other stipulations in even the most basic licensing agreements suggest that economists have taken an overly simplistic view of the situation.\textsuperscript{33} Lump-sum charges associated with sliding rates are routine in licensing agreements, together with various use restrictions and terms. Lump-sum extractions are indeed employed to encourage marginal uses.\textsuperscript{34}

This makes somewhat mystifying the frequent use of royalty rates in patent or trade-secret licenses. Not only will these inhibit marginal use, but the costs of enforcing a stipulated royalty payment would seem to be higher than a lump-sum charge because a royalty necessitates the metering or checking of output units. Why not let lump-sum (or annual) charges substitute for royalty rates? My explanation is that the licensor

\textsuperscript{33} This judgement is based on an examination of about 80 patent licenses.

\textsuperscript{34} See Plus Promotions, Inc. v. RCA Mfg. Co., Inc., 49 F. Supp. 116 (S.D.N.Y. 1943), an unusual case involving an idea submitted to RCA without a contractual relationship. Suggestion was made that instead of paying royalties to famous artists to produce records, the company should pay them an agreed sum and then sell the records at lower costs. RCA apparently adopted the idea, but the court denied recovery on grounds that the agreement was not sufficiently concrete. What is remarkable about this case is not the idea itself, nor that the method of pricing a good may be subject to trade-secret protection, but that RCA had not previously known about this pricing method!
of the patent does not know how much lump-sum to charge. This is particularly true in the face of changing market conditions or when costs are high for obtaining information on how well the market will receive the invention. The royalty rates serve as a useful metering device; without this channel of information, the licensor is insulated from the market by the licensee, who will tend to understate the market value of the invention. The imposition of unit royalty rates and the implied right of checking output by the licensor yield useful information to approximate the lump-sum charge. This, I surmise, is why lump-sum charges are often used in association with unit rates structures. Furthermore, for both the licensor and potential additional licensees the use of unit rates also provide useful signals for the issuance of further licenses in other market areas and for the revision of contractual terms in license renewals.

In addition to uncertainty and indivisibility, Arrow considers that the inappropriability of returns to invention also leads to underinvestment in innovation. He notes two characteristics of invention which impede appropriability. One is that "the very use of the [invented knowledge] in any productive way is bound to reveal it, at least in part" and that "its value for the purchaser is not known until he has the [knowledge], but then he has in effect acquired it without cost" -- hence, the difficulty of fully exacting payment.\(^{35}\) In some dimensions, this argument

\(^{35}\) Arrow "Economic Welfare," p. 615.
is correct. However, it is difficult to enforce ownership over an unknown. Disclosure is a prerequisite even for the drafting of patent claims, and capturability is often enhanced by the protection granted after the disclosure. 36

The second appropriability problem cited by Arrow hinges on his distinction between using an idea for further research and using an idea in producing commodities. He argues that the capturability problem is more severe with the former:

To appropriate information for use as a basis for further research is much more difficult than to appropriate it for use in producing commodities; and the value of information for use in developing further information is much more conjectural than the value of its use in production and therefore much more likely to be underestimated. Consequently, if a price is charged for the information, the demand is even more likely to be suboptimal. Thus basic research, the output of which is only used as an informational input into other inventive activities, is especially unlikely to be rewarded. In fact, it is likely to be of commercial value to the firm undertaking it only if other firms are prevented from using the information obtained. But such restriction on the transmittal of information will reduce the efficiency of inventive activity in general and will therefore reduce its quantity also. 37

By the last statement, Arrow applies to basic research his earlier point: namely, that if a price is charged for use of the information, the level of utilization will be "suboptimal." Beyond that, he offers two reasons for the lesser appropriability of returns to basic research. One is that the more basic the information, the more uncertain will be its market

---


value to a buyer averse to risk.\textsuperscript{38} Second, the difficulty of preventing others from improving upon a basic invention implies greater difficulty in exacting payment. The latter point is decidedly controversial and deserves fuller analysis.

III. Basic Research: The Issue of a Blocking Patent

Whereas Plant argued that a patent granted to basic research will inefficiently block the further development of ideas, Arrow's view is that, ignoring marginal-cost pricing, such a patent grant lacks sufficient blocking power. That is to say, while Plant's argument implies that patents overreward basic research (excessive appropriability), Arrow argues that even under patent laws basic research is bound to be underrewarded. And whereas Arrow views the patent system as underrewarding basic research (inadequate appropriability), Knight had earlier gone so far as to say that the system underrewards the basic researcher: "It is undoubtedly a very rare and exceptional case where the really deserving inventor gets anything like a fair reward."\textsuperscript{39}

According to Knight, because of risk and the problem of inappropriability, the system chiefly rewards the one who puts on the finishing touch. It is strange that, as a major innovator of the meaning of competition, Knight would have advanced this view. Even if the latter argument holds,

\textsuperscript{38} Arrow, however, did not assume general risk aversion, and a preference for risk, which Arrow allowed earlier in the paper, may lead to "excessive" basic research.

the patent system does not penalize the true inventor, since he has voluntarily relinquished his option to be the finishing toucher himself. And competition will insure that, pecuniary and nonpecuniary aspects considered, the reward to him will be as "fair" as to anyone else. A windfall gain for the finishing touch must be offset by a windfall loss elsewhere.

So the questions remain: Does a patent system underreward basic research? Or do patents overreward basic research and thereby block efficient improvements? In framing the answers, let me refer briefly to the patent system in the United States. We will see that the replies lie somewhere between the extremes voiced by Plant, Arrow, and Knight. To varying degrees, each of their positions -- and indeed the major controversy over "blocking patents" in itself -- reflects an inaccurate view of the relationship between (1) the right to use an idea to produce another idea and (2) the right to use an idea to produce a commodity.

The term "blocking patent" has been a confusing one.\textsuperscript{40} Strictly speaking, every effective patent blocks, for the right to block means the right to exclude. In a trivial sense, a royalty imposed on the use of a patent for production is one result of the right to block. The matter becomes complicated, however, when the patent by itself generates no marketable commodity or when a patented commodity is suppressed partly

or wholly by a superior one, for the question then follows: What valuable right does the original patentee have if someone tries to improve upon his ideas? It is in this context that the concept of "blocking patent" is usually employed.

To facilitate discussion, let us call the right to use an idea to develop another idea a "development right," and the right to use an idea to produce a commodity a "production right." Historically, patent laws have granted both rights to the patentee. However, evolution of U.S. patent laws has shown that direct protection of development rights generates ambiguity in the law.\footnote{When the U.S. patent system was introduced in 1793, the patent law attempted to protect development rights directly [c. 11, 1 Stat. 318]. And it was not until 1836 that patent claims began to play a significant role: "[the applicant] shall particularly specify and point out the part, improvement, or combination, which he claims as his own invention or discovery" [c. 357, 5 Stat. 117]. The protection of claims was further strengthened under the Act of 1870, which required that the patent applicant "shall particularly point out and distinctly claim the part, improvement, or combination which he claims as his invention or discovery" [c. 230, 16 Stat. 198]. (Italics added.) In short, the early effort to protect development rights had left the U.S. system in such a chaotic state that patent rights were virtually unenforceable. See P. J. Federico, "Commentary on the New Patent Act," in United States Code Annotated, Title 35 -- Patents, Section 1 to 110 (St. Paul : West Publishing Co., 1974), pp. 2-5.} Now the market value of any patent must ultimately rest on some marketable product actually or potentially to be generated under its provisions. The development rights of a patent are not directly enforceable or protectable, in the sense that there is no way a patentee can exclude others from thinking about his disclosed idea or tinkering with it in a garage. However, development rights are often indirectly protected by the enforceable exclusionary right to turn out
a product. And because any infringement of production rights can be traced to a market transaction, the exclusive right to produce an observable product implies exclusive rights to improve upon a patented idea. Let me extend that explanation.

Suppose an inventor comes up with a patentable idea which has no immediate commercial value. When he obtains a patent, his claims are disclosed; subsequently, regardless of how the patent law is written, he has no way of preventing anyone else from thinking about the revealed idea or from experimenting to improve the original. But suppose the experimenter then tries to market his improvement.

First, let us assume that an original patent covers an article consisting of Parts A and B. Inspired by the disclosure, the experimenter comes up with a better article serving the same purpose but consisting solely of Parts C and D. He then obtains a different patent and proceeds to market his product. An infringement suit will not benefit the first patentee, since his claim in no way dominates the new product. It does not follow, however, that his patent becomes worthless: it may yet be incorporated in some other future designs.

Next, suppose the experimenter has improved the original patent by adding Part C, and for this addition he obtains a patent. If he produces and sells only Part C (which has value only if used in conjunction with

---

42 In the United States, it would be difficult to prevent someone from soliciting orders in advance to produce a patented article as soon as the patent expires. See Berle and de Camp, *Inventions, Patents* ...., p. 362.
Parts A and B) he may or may not be subject to infringement charges. If the court rules in the original inventor's favor, then the protection of his production right in some degree implies the protection of his development right.

Finally, suppose that the improvement implicit in Part C is inseparable from Parts A and B. Then in trying to sell a product consisting of Parts A, B and C he will clearly be dominated by the old patent, and only by license from the earlier inventor can he expect to avoid litigation. However, tenuous may be the original inventor's plans for production, he exercises as much control over the development of improvements as over production rights -- provided that the production rights are effectively protected.

In short, the rule for improvement seems to read: You may tinker with my patent any way you please, but plan to pay me when you produce any commodity over which I have some claim; moreover, to avoid my possibly excessive demands, it may be wise for you to obtain a license from me in advance. Thus a patent that confers production rights on the patentee does imply some exclusive rights on development but only to the extent that the improvement is at least in part dominated by the original invention. The blocking effect so implied enhances the appropriability of return to basic research. On the other hand, a patent grant of production rights

43 In the U.S., if it can be shown that the seller of Part C knew that this part would haveto be used with Parts A and B, and that a buyer so using it would infringe the original patent, then the seller of Part C may be charged with contributory infringement.
does not prevent anyone from thinking about the patented idea or using it to produce a different, improved product not embodying the original invention. To this degree the patent grant lacks blocking power and encourages imitation. 44

In all consistency, one cannot advocate a patent system on grounds that it encourages new products while at the same time totally denying its blocking effects on improvement. However, other factors further mitigate any blocking effects that might otherwise flow from patent protection of basic research. First, the issue frequently is moot simply because the improvements on the patented idea are made by the original inventor himself, and he enjoys the comparative advantage of prior experience. 45

44 In actual practice, of course, court rulings on infringements are often arbitrary and even contradictory. In Wright Co. v. Herring-Curtiss et al. [204 F. R. 596], a case involving airplane wing design, the ruling in Wright's favor rested on a broad interpretation of the original patent, although scarcely any overlapping claims were detectable. In the Selden automobile engine patent case [Columbia Motor Car Co. v. C. A. Duerr and Co., 184 F.R. 893], a debatably narrow interpretation of the original patent defeated what had been a consistent winner. Even in the classic Edison Lamp case in which the court upheld Edison's basic patent twenty-two times [see Berle and de Camp, Inventions, Patents ..., pp. 104-7], Edison spent more in legal fees than he received from his patents. [Goller, Competing ... Patents, pp. 727-8]. In all these infringement cases involving improvements, as in other similar cases, the protection problem stems from ambiguities inherent in the production rights which led to ambiguity in development rights. See D. R. Dunner, J. B. Gambrell, and I. Kayton, Patent Law Perspectives (New York: Matthew Bender, 1974), Vol. 2, Section B.1.

45 Even ignoring discounting considerations, a patent lawyer informed me that the seventeen-year term of a patent grant is often not a severe restraint on the inventor, because before the patent expires it may have been rendered obsolete by the same firm.
Second, in the United States the patent law specifies that a "pure" idea is not patentable, precluding any blocking power that might occur if theoretical principles discovered by basic research were eligible for patent protection. Specifically, American patent law requires not only novelty and "non-obviousness" as criteria of patentability, but also utility -- a standard based on marketability. To forestall the possibility of espionage or independent discovery by others, an inventor having no immediate plans for production may seek patent protection along the way of his development process. If the criterion of utility was to be tightened, fewer basic ideas would be protectable by patent; but the overall effects of such an action could not be assessed without a comprehensive study of such other protective alternative as copyrights, trademarks, and trade secret laws.

On the other hand, in the eventuality that an original patent actually exerts effective blocking power, the obstacle to improvements may be reduced by any of several contractual arrangements. For example, ordinary patent licenses seldom restrict improvements by the licensee. Patent assignment is an outright transaction which even more evidently cancels the blocking action, leaving the purchaser free to make improvements at his own pace; numerous patents that are never directly used in production have been transacted from this purpose. Another common device is cross-licensing or patent pooling, where each participant contributes some

related patent, often on a royalty-free basis, to clear the way for all. Such agreements routinely provide immunity from litigation, because the high costs of infringement suits (due in particular to the uncertainty of court interpretation of claims) would greatly reduce the value of the patents. Pooling arrangements to eliminate improvement blocking -- and to avoid litigation -- are recognized by the courts:

An interchange of patent rights ... is frequently necessary if technical advancement is not to be blocked by threatened litigation. This is often the case where patents covering improvements of a basic process, owned by one manufacturer, are granted to another. A patent may be rendered quite useless, or "blocked," by another unexpired patent which covers a vitally related feature of the manufacturing process. Unless some agreement can be reached, the parties are hampered and exposed to litigation. And, frequently, the cost of litigation to a patentee is greater than the value of a patent for a minor improvement.

47 The distinction between a patent pool and a cross-license is that with the former the patents are usually held in trust by one agent who issues licenses to all members of the pool. This is the arrangement usually chosen when many patents are involved.

48 There are other reasons for pooling patents than to eliminate improvement blocking, but a detailed discussion of patent pools and cross-licenses is beyond the scope of this paper. My reference to these forms of contractual arrangements is drawn not only from the cross-licenses I have at hand, but also from the extremely valuable information to be found in House Hearings on the Pooling of Patents [Hearings on H.R. 4523] (Washington, D.C.: U.S. Government Printing Office, 1936) Pts. 1-2.

49 In 1959, Berle and de Camp estimated that the average cost per case for one side of a litigation was between $10,000 and $20,000; Inventions, Patents ..., p. 367.

50 Standard Oil Co. v. U.S., 283 U.S. 163 (1931) at 171.
The above discussion underscores the futility of attempts to estimate patent utilization. Past studies have claimed that somewhere between 40 and 80 percent of the patents in the United States have never been put to use. But does this imply a remarkable lack of foresight which led inventors to overinvest in basic research? A patent may be obtained, assigned, or pooled for the development of future products. There is no royalty. Yet when a commodity finally emerges, several hundred patents may have helped along the line. Even in an ordinary patent license where a royalty rate is clearly expressed, it is not uncommon that a host of patented devices may be leased without specific reference, or that a single patent may be specified with a package of other knowledge thrown in "free" of charge. Now how is the director of a research and development laboratory supposed to answer a questionnaire from some investigator asking him what fraction of his patents are used in production or have earned a royalty?  

51 For an exposition of the various attempts which have been made to determine the "use-rate" of patents, see Normal J. Gharrity, "The Use and Non-Use of Patented Inventions" (Ph.D. dissertation, Johns Hopkins University, 1965). For various estimates of the rate of utilization, including those calculated by Gharrity, see Table IX-1, p. 265.  

52 The discussion here does not imply that all patents are used! For evidence, see A. E. Brown and H. A. Heffcott, Absolutely Mad Inventions (New York: Dover Publications, Inc., 1970).
IV. Concluding Remarks

Without doubt the patent system as we know it today is a highly costly process, if only in its potential for litigation. Ideas are hard to sell. My contention is that the chief factor inhibiting the capture of returns to innovation is the exceptionally high cost of transactions associated with protecting the property rights to an idea — originating with the great difficulty of defining and delineating the idea itself.

It is far from easy to identify the costs of delineating and transacting ideas, and the task is further complicated when other systems of protection, such as trademark, copyright, and trade secret, must be considered jointly with the patent system. We have attempted here to spell out some of the relevant issues and to point out where past writings have fallen short of reaching satisfactory answers.

We have gained much in a systematic study of these earlier analyses of an important and challenging subject. From wholly disparate vantage points, Bentham, Say, Mill, Clark, Taussig, Pigou, Knight, Plant, Arrow, and others have provided useful, if contradictory, insights into property rights in invention. All have added to our understanding of the effects of a patent system and, in my view, their honors in the literature are richly deserved.

My disagreement with Plant and Arrow stems largely from our differing emphases on transaction costs (which are implicit with Plant, explicit with Arrow) and which lead to increasingly divergent implications the farther one proceeds. In my view, however, the works of Plant and Arrow stand almost alone in shaping current thinking on the economics of invention.
Plant's concern with monopoly pricing is central to the present-day debate on the compatibility of patent and antitrust laws, and another point briefly noted by him has gained stature as the "single-prize thesis." This is his argument that the patent system renders "almost nugatory the labours of all the rest." In different contexts, Knight had earlier hinted at the thesis of the dissipation of rent and H. Scott Gordon followed with an exposition of such "dissipation" in marine fishery. The current thesis of a "rush to invent" and of "wasteful duplication of invention" owes much to these works.

---

53 See, in particular, Bowman, Patent and Antitrust Law.


Finally, the work of Arrow may be classed as monumental if only for his treatment of the problem of appropriability. The same problem was recognized by Pigou and Knight earlier, but Arrow brings it to bear on the various interesting attributes of ideas or information. The transaction-cost paradigm has just begun, and investigation into the physical attributes of resources has proved to be a most fruitful approach to ascertain the variation in transaction costs as a basis for the derivation of testable implications.
Patents: Monopolies or Property Rights?  
Edmund W. Kitch

The patent system has been viewed by economists as a trade-off between the incentives created for the production of inventions and the losses imposed by the temporary monopoly the patent confers. "Stimulating the invention and development of new products and processes is without doubt the most important benefit expected of the patent system. For it society pays a price; the monopoly power conferred by patent grants." In a paper which my commentator for today has characterized as "little influenced by any concern for reality," and which two other members of this conference considered of sufficient importance to merit a thoughtful comment, I argued that the traditional view failed to consider the positive effects which a patent system has on the output of new inventions from the information embodied in old inventions. Here, I will discuss the other half of the traditional position, that patents are monopolies which impose a social loss by constricting the use of the patented invention.

Those of you familiar with the long policy debates over the patent system will recognize that I am taking up a position long argued by the patent lawyers. To quote the much garbled current edition of a leading 19th century treatise on patents, "a patent grant should never be referred to as a monopoly—even a legal monopoly. Since a monopoly was a crime under common law and is one under statute, it is evident that the terms should not be
applied to a legal transaction as evidence by a legal instrument or deed, to wit, a patent grant. In considering patent matters and questions, letters patent should be viewed in the same light as a land grant, should never be termed a monopoly, should be accorded the same treatment as other forms of property, and should always be referred to or designated as a patent grant. No wonder that under the influence of such persuasive argumentation, I have long considered it obvious that a patent is a monopoly.

The propositions that I wish to argue here are first, that a monopoly should be defined as a condition in which a seller faces a demand curve with a negative slope, and second, that there is no valid a priori reason to assume that most or even many owners of patents are monopolists in this sense. Whether, as a matter of empirical fact, some or even many owners of patents face demand curves with a negative slope is an issue that has never been seriously investigated, probably because it has been generally assumed that the answer is clear, but casual analysis of one striking, litigated antitrust case suggests that it may be a rare phenomenon indeed.

The issue of how monopoly should be defined is one best left to lexicographers or to the high priesthood of economic terminology, a priesthood for which I am disqualified by training in spite of of my fellow-traveler membership in the American Economics Association. Certainly common usage, and not infrequently the usage of economists, gives the term "monopoly" multiple meanings. For instance, if "monopoly" is a synonym for an exclu-
sive right, then it is true by definition that the owner of a patent is a monopolist. But I will not pause to argue at length to an audience of the economic sophistication of this one that the meaning of monopoly which is of concern to the scientific student of economic behavior is that condition which generates social loss, a condition which is present only when the demand curve has a negative slope.

Most would surely concede, at least upon reflection, that in this sense of monopoly, it is clear that not every owner of a patent has a monopoly. Many patents issue upon technologically trivial or commercially unimportant inventions (which are not necessarily the same things) that have numerous close substitutes. Polaroid's technologically splendid and patented instant movie system was felled by competition from video cameras. A patent on one fastening system may simply provide an alternative to numerous other fastening systems. And so on. The area of dispute, I think, would be over whether the patent as monopoly is an important case, which occurs frequently. I also assume that numbers of patents are an unimportant measure, and even if a small percentage of issued patents are monopolies, this still might be the most important class of patents because they would account for the largest part of production occurring under patent rights. I wish to assert that the ownership of patents is no different than the ownership of any other property right necessary as an input, and that we should no more assume that the owner of a patent is a monopolist than we should assume that the owner
of particularly fertile land, especially productive skills, or of an advantageous location is a monopolist.

The issue is seldom discussed except in passing, and often the term "monopoly" is used ambiguously, either in the sense of an exclusive right, or in the sense that the owner of the patent faces a negatively sloping demand curve. Not unrepresentative is Machlup's classic discussion. Reviewing the literature, he tells us that "While some economists before 1873 were anxious to deny that patents conferred 'monopolies' and, indeed, had talked of 'property in inventions' chiefly in order to avoid using the unpopular word 'monopoly'--most of this squeamishness has disappeared." But a few pages later, he asserts confidently that "Patents, by giving their owners exclusive rights to the commercial exploitation of inventions, secure to these owners profits (so-called 'quasi-rents') which are ultimately collected from consumers as part of the price for goods and services." Bowman, who uses more rigorous price theory, is less ambiguous. "The temporary monopoly afforded by a patent, once a particular invention has come into being, will have all the output-restrictive disabilities of any monopoly."

I believe that I have two insights of some novelty to offer to the analysis of this issue. The first relates to the importance of the newly obsolete technology in the early years of the patent as competition for the patented technology. The second relates to the incentives operating on the patent owner as the end of the patent term approaches to increase output in order to "occupy" the market when the patent expires.
I have elsewhere pointed out that the patent owner's commercial product will often reach the market a substantial number of years after the critical patents have issued, for reasons inherent in the structure of the patent system. When the product is ready for market, however, it is the rare invention indeed which immediately sweeps aside its predecessors. Transistors did not displace vacuum tubes overnight, word processors did not immediately eliminate typewriters. The reason is that there are substantial investments committed to the old technology. These are not only investments in special purpose machinery, whose economic value may have fallen to zero, but also investments in specialized human capital which can be used only with the old technology. For the user, the cost of new technology is not only the price, but also the investment required to enable him or his personnel to make use of it. Thus even obsolete technology may present very real competition to the new technology that has made it obsolete. Obsolescence only means that buyers will not make additional, new investments specialized to the old technology. It does not mean that they will immediately stop using it.

Potential competitive entry by other firms once the patent expires, or by other firms with competitive technology not within the scope of the patent claim, exerts an important force on the market for the products subject to the patent even before the patent expires. During the life of the patent, the patent owner can attempt to develop the specialized resources necessary to serve the market for the patented product that will exist when
that market is open to entry by others, to occupy the field so that the expiration of the patent will not induce other firms to duplicate his specialized investments to take advantage of the larger market revealed by the lower post patent price. But the only way he can do that is, prior to the expiration of the patent, serve those customers who will be customers for the product once the patent expires. And the only way he can identify and prepare to serve that market is to lower the price down to the competitive price. Put another way, a marginal cost of pricing up during the life of the patent is the loss of market position after the patent expires. Only if the patent owner can instantaneously and costlessly expand his market the moment the patent expires is this force unimportant.

This paper first analyzes the issue in terms of formal price theory, and then illustrates the analysis with a discussion of how the issue has been handled in American antitrust cases, with special focus on the antitrust litigation involving the Xerox patents, an example of patents that many would view as a clear case of monopoly.
Price Theory and the Patent

The way in which most economists analyze this problem is through the example of a cost reducing invention. They diagram it this way:

![Graph showing cost and demand lines]

Figure 1

This figure, of course, does not portray the market for the patent, but the market for the product produced through the use of the patent. I think most analysts view that as a trivial point, and would be quick to draw the following diagram for the derived demand for the patent:
However, I have not actually found an example of this figure in the literature, perhaps because it has an obvious defect. The defect is that it is impossible to determine what the scale for the X-axis is. In the diagram for the cost-reducing invention, the X-axis was units of the product produced by the use of the cost-reducing invention. One can get from that diagram to the patent diagram only if one assumes that the patent is used in fixed proportions, that one "unit" of the patent is used with each "unit" of the end product. I think the unexamined assumption that fixed proportions are a reasonable assumption accounts for the fact that analysts have assumed that the end product diagram captures the problem when it does not.
There is no persuasive reason to assume that the information embodied in the patent is utilized in the end product in any fixed proportion. The apparent frequency of royalty schemes that charge a fixed price per unit may create the impression that this is so, but a convenient method for metering the use of the patent does not really reveal what the economic proportions are. Take as an example a patent for the design of a machine. Suppose that the machine is sold in two models, one costing $10,000 and one costing $2,000? Is the patent used in the same amount in the $10,000 machine as in the $2,000 machine? Is the difference to be accounted for by other inputs, or by the fact that one machine "uses" the patent more than the other?

The patent owner, of course, can optimize the value of his patent if he can charge more for it where its contribution to the value of the final product is greater, and less where the patent contributes less. A patent owner has one very simple method of doing this, and that is to exploit the patent at least in part himself. He can then allocate the cost of the patent to the various uses he makes of it in the proportions that optimize the value of the patent. The point of this analytic agnosticism is to suggest that there can be no appropriate price-theory analysis of the patent itself. A patent is a single, unique thing that must be exploited at a particular point in time. It might be diagrammed like this:
Figure 3 shows a single patent, sold at a single price. That transaction can give us no information about either a demand or supply schedule for that single patent.

The important point is that it is impossible to analyze a market for the patent. It is only possible to analyze a market for a product that embodies the patent. The problem with Figure 1 as an analysis of the problem is its assumption that marginal cost is constant, that the firm experiences no diseconomies of scale until it reaches a region of output in which the demand curve slopes downward. We commonly make such an assumption in the natural monopoly case, but there is no reason to assume that the firm manufacturing the product that uses the patent has a natural monopoly in its production. The implicit story that
motivates the assumption of fixed marginal cost is the assumption that the patent is used in fixed proportions through all level of output, and the only significant cost that might vary is the cost of using the patent itself. But patents account for a very small part of the costs of most products that embody the patent. Why shouldn’t we assume, as we usually do, that the costs of those inputs begin to rise before the demand curve slopes downward?

An important source of confusion in the analysis of these issues is the following question: How can a patent have any value if it is not a monopoly. If the patent owner faces competition and his marginal cost of using the patent is zero, then why wouldn’t competition mean that the value of the patent is zero, since under competition price equals marginal cost? The answer is that the patent can affect the patent owner’s (or the licensee’s) average cost of production, but not its marginal cost. A patent can have value like any input that gives a firm a comparative advantage over its competitors, but that does not mean that the owner of the patent owns a "monopoly."

There are addition reasons for thinking that it is not unlikely that the nonpatent costs associated with the manufacture
and sale of a product which uses patented information will cause production to occur in regions where the demand curve is not downward sloping. In the early years of the patent, the patent owner will face all of the costs associated with the introduction of a new technology. Although the technology being displaced may be "obsolete" in an engineering sense, it will often remain as a competitive force for many years, and be priced to reflect the obsolescence of the specialized resources dedicated to its production and use. Then as the patented technology begins to dominate the market, the patent owner must consider the coming end of the patent term, and the ever present possibility that some newer technology may appear to displace his patent. His patent gives him the opportunity to occupy the field, and if he he able to develop an organization that can manufacture and sell the patented product at costs equal to any potential entrant, he may occupy it for many years. But he can only occupy the field if he prices the product at a level which identifies and develops the market that exists at price levels close to cost. That is why many firms with innovative products can be seen to rush forward in their effort to introduce them as fast and as widely as possible. The strategy of the output-restraining monopolist will often be costly in the long run.

These arguments do not mean that no patents confer monopoly power upon their owners. Some of the drug patents for instance, patents which achieved dramatic reductions in the cost and effectiveness of medical care, probably have conferred monopolies upon their owners. But this analysis suggests that they are very
exceptional cases.

Patents in the American Courts: Are they Monopolies?

American courts have faced the issue of whether patents are monopolies in two quite different contexts. They have faced the question, like economists, as an issue of theory in the process of formulating antitrust rules based upon the economic effects of patents. And they have faced the question, unlike economists, as a factual issue. Whether a particular patent in a particular market conferred a monopoly upon its owner. In the first context the courts have answered, like the economists, that a patent is a monopoly. In the second context, they have answered, either expressly or by necessary implication, that a patent is not a monopoly.

The courts first faced the problem of whether a patent is properly viewed as a monopoly in the context of tie-in sales. These were cases where the patent owner had licensed the patent on condition that the licensee use unpatented products to be supplied by the licensor. The cases asked the question whether such a contractual arrangement violated the antitrust laws on their face. Without any record evidence as to whether or not the patents before the court were or were not monopolies, the courts held that such contractual arrangements did violate the antitrust laws because they extended the patent monopoly to unpatented products. Once this law had been established, the courts then faced the problem whether a contractual provision which tied any
other product or service to a property right violated the anti-
trust laws. The defendants argued that it did not because other
property rights, unlike patents, were not monopolies. But the
Supreme Court held that for this purpose they were the same, a
view entirely consistent with the analysis here.

The courts have also addressed the question of whether a patent is a monopoly in a procedural context where the proposi-
tion had to be proven as a matter of fact. For instance, it has
been held to be a violation of the antitrust laws to obtain a
patent by fraud if the patent conferred a monopoly in fact.
These cases have been defended successfully on the ground that
the particular patent in issue did not confer a monopoly. Simi-
lar results have been reached in other litigation about patents
and in litigation involving other industrial property rights.

The case of SCM Corp. v. Xerox Corp., 645 F.2d 1195 (2d Cir.
1981), supplies a dramatic example of a litigated case in which
the courts resisted the conclusion that the patents there involv-
ed had conferred a monopoly. The case was an action under the
Sherman and Clayton Acts by a competitor of Xerox for damages
caused it by Xerox’s allegedly illegal monopolization of the
office copier industry. SCM lost the case in spite of the fact
that it had a facially strong claim.

SCM manufactured and marketed an office copier that required
the use of specially prepared, coated paper. Xerox entered this
market in 1960 with a copier that used plain paper and by the end
of the decade it dominated the by then enormously expanded mar-

14
ket. SCM sought as damages (trebled of course) the money it could have made had Xerox let SCM use Xerox's patents for a plain paper copier.

That may strike you as an absurd antitrust claim. At least, reading between the lines, it must have struck Judge Newman, the (then) District (now Circuit) judge who tried the case, and the panel of the Second Circuit who affirmed his decision, that way. If Xerox had simply developed and patented xerographic technology, and then used that technology to successfully build and market a plain paper office copier, the antitrust issues might have been quickly disposed of. The purpose of a patent is to confer an exclusive right to make and vend the patented product, and if the patents had been properly obtained, then they would only have had the effect they were intended by Congress to have.

Unfortunately for Xerox, its entry into the office copier market had not occurred so simply. Before its antitrust troubles were over, it had entered into a consent decree with the Federal Trade Commission requiring it to license its patents at nominal royalties and had experienced the return of a jury verdict holding that it had violated the antitrust laws and caused damage to SCM in the amount of $37.1 million, or $111.3 million trebled. All that stood between Xerox and a diminution of its net worth was the judges.

The technology of xerography was originally discovered -- perhaps imagined is a better word-- by on Chester Carlson, he hit upon the concept of making images by using photoconductive sur-
faces. His was a major achievement, a method of image reproduction entirely different from photography. He used the process to make crude images in 1938 and obtained patents. It was, however, to be a long haul from the concept to the commercial product. After extended search for financial backing to develop the invention, he licensed his patents to Battelle memorial Institute, a non-profit research corporation in 1944. Battelle further developed the technology and searched further for financial backing. In 1946 it entered into an agreement with the Haloid Corporation of Rochester, New York (later to become Xerox) for support of its continuing research on xerography. Under the agreement, Battelle gave Xerox (then Haloid) a non-exclusive license on its patents. Battelle and Xerox pursued development of the technology further. By 1956 Xerox had succeeded in marketing some specialized but profitable products using xerography and believed that it knew how to make an office copier. In that year, I entered into a new agreement with Batelle that gave Battelle a stock interest in Xerox for an exclusive license under the patents. This agreement effectively merged Battelle's and Xerox's patents and technological know-how, and insured that Battelle would not use its patent position to create other competitors in the plain-paper-copying field. This agreement was the heart of SCM's antitrust complaint.

The argument that the agreement violated the antitrust laws was straightforward. The 1956 agreement was an acquisition of assets within section 7 of the Clayton Act. Since the effect of the agreement was to foreclose any possible competitor in the
production of an office copier using xerography, it was an acquisition of assets whose probable effect was to tend to create a monopoly in plain paper copying. Q.E.D., as a logician might say.

The case was tried in an unusual and very successful manner. It was tried in three phases, one for liability and two for elements of damage. Along with the general verdicts, the jury was given numerous special interrogatories designed to prevent the need for a retrial should any of the instructions prove erroneous. The resulting verdicts of the jury were internally consistent and consistent with the evidence presented. The 1956 agreement, the jury found, had tended to create a monopoly and had caused injury to SCM. The jury also found that SCM should have reduced its damages, by suing at an earlier date to force Xerox to grant a license, a charming if somewhat naive vote of confidence in the efficacy of preventive antitrust relief in the federal courts. The judges did not choose to rely on that finding in reaching a no liability result.

Judge Newman responded to these jury verdicts by concluding that they were consistent with the evidence and the instructions, but he would not enter a judgment for damages because such a result would be inconsistent with the patent laws. "Liability for retrospective money damages cannot be predicated under sec. 7 upon a patent acquisition made prior to the existence of a relevant product market. No case has ever upheld such liability, and its allowance in this case would be beyond any reasonable accom-
modation of the patent and anti-trust laws. Indeed, there is considerable doubt whether liability can be grounded under sec. 7 for the mere acquisition of any asset prior to the existence of a relevant product market." 463 F. Supp. 1080. His discomfort with this approach is suggested by the fact that he did not hold that the transaction was not a violation of sec. 7, only that SCM could not collect damages for the violation.

The Second Circuit affirmed even more boldly. "[N]otwithstanding the jury's implicit finding [that the 1956 agreement had the reasonably foreseeable effect of giving Xerox an eventual monopoly], we conclude that under the facts presented here, the policies of the patent laws preclude the imposition of antitrust liability." 645 F.2d 1209. In a diffuse and confused discussion, the court concluded that neither sections 1 or 2 of the Sherman Act nor section 7 of the Clayton Act could be violated by an agreement that restrained trade in a market that did not exist at the time of the agreement. Since a plain paper copying market did not exist in 1956, an agreement monopolizing that market in 1956 could not violate the antitrust laws. "[I]rrespective of the jury's implicit finding that Xerox's commercial success was reasonably foreseeable in 1956, Xerox was lawfully entitled to purchase the patents it did pursuant to the agreement it made with Battelle that year." 645 F.2d 1209.

The suggestion implicit in the logic of the Second Circuit that any patent agreement that relates to technology not yet commercialized is legal under the antitrust laws is startling.
given the focus of section 7 on the potential future effects of the agreements within its terms.

The rest of this paper will discuss the central antitrust issue—whether or not the 1956 agreement really had a foreseeable tendency to confer monopoly power on Xerox. However, there were two other important problems with SCM's antitrust claim that should be noted.

First, in a functional sense it doesn't make much sense to think of the 1956 agreement as involving an agreement between different firms. The research to develop xerography had been carried on by Battelle and Xerox since 1946 as a kind of joint venture, and they were for purposes of that venture really a single firm. The 1956 agreement was more like an agreement in which rights to an invention are rearranged among co-participants in light of the inventive results and marketing possibilities than a joining together of potentially competitive firms. Any firm that holds and decides to exploit internally a patent is deciding not to license others. The 1956 agreement was a formal way for the joint venture to determine that its best interest lay in internal development and exploitation. The Second Circuit was clearly bothered by the idea that Battelle and Xerox together could not do what either could have done separately had one of them developed the invention alone. A limited way to reach the result of no liability would have been to adopt a functional, economic concept of firm and to conclude that since there was no agreement between firms there was no agreement involving sec. 7 at issue. Put another way, the 1956 agreement was not a merger of
potentially competing patents by separate firms. Antitrust law, however, has had trouble treating formally separate firms as single enterprises, as the intra-enterprise conspiracy cases show.

Second, SCM had serious problems with its damage theory. If the 1956 agreement restrained competition between Xerox and potential competitors who might have been licensed by Battelle, then the persons injured by the agreement would be purchasers of copying machines who might have been able to purchase them more cheaply had Xerox faced more competition. SCM’s theory was that it was damaged because it had lost the profit opportunities it would have had if it had been licensed. Assuming that SCM could make a plausible showing that if licenses had been offered it would have obtained one, it still faced the problem that the royalty charged for the license should have been equal, ex ante, to the profit opportunity conferred by the license.

In SCM Corp. v. Xerox Corp., the District Court put the question of whether the patent agreement had tended to create a monopoly to the jury, and the jury answered it as follows (463 F. Supp. 1022):

20. Was the probable effect of Xerox’s acquisition of patents pursuant to the 1956 Xerox-Battelle agreement, when the agreement was made, substantially to lessen competition or to tend to create a monopoly in any relevant market or sub-market that you have found to exist?

<Check whatever market or sub-market, if any, you have found
to exist and then answer this question only as to such market or sub-market, or both):

   X  In a market consisting only of "convenience office copiers" using both plain and coated paper?

       Yes X
       No   ___

   X  market or sub-market consisting only of "convenience office copiers" using only plain paper?

       Yes X
       No   ___

The jury's answer was eminently sensible. The 1956 agreement insured that only Xerox would have the legal power to manufacture and market a copier using xerographic technology. The superiority of that technology was sufficiently likely in 1956 that it was foreseeable that Xerox would end up as the only seller of convenience office copiers. Whether or not that state of the facts should result in antitrust liability is a separate question.

The problem for the courts in SCM Corp. v. Xerox Corp. began with the concept of relevant market used in the case. The court used the concept, common in the antitrust cases, that there is to be a "relevant market," and that a "relevant market" is an all or nothing proposition. The jury found that plain paper copiers were a relevant market, and Xerox manifestly being the only seller in that market, it was difficult for the courts to escape from the conclusion that the market had been monopolized. But markets are matters of degree, and depending upon such thing
as supply substitution and demand substitution, a seller in a single market may or may not have monopoly power. The jury had also found that there was no relevant market in plain paper copiers in 1964, a finding that at that time competition from other forms of copying meant that plain paper copiers were not an economic monopoly. That is in no way inconsistent with its finding that the agreement in 1956 had the reasonably foreseeable effect of creating a monopoly in office convenience copiers in the future, and the jury found that such a monopoly did exist by 1969.

The key issue in the case, however, was whether it was likely that the 1956 agreement would tend to create a monopoly. In the terminology used here, the issue would be whether Xerox in 1956 should have expected to face a negatively sloping demand curve for its product because of the protection provided by its patent agreement with Battelle. Working from the facts as they are reported in the opinions, it seems to me that that was very unlikely. The reason is that Xerox continued to face competition from the then existing stock of technology, and from technology still to be introduced. First, the sellers of the competitive technology continued to manufacture and market products, and would continue to do so until the price they could obtained fell below their marginal cost. And even after sellers of new equipment left the market, that would not end the competition from the old technology. Owners of existing machines would continue to sue them until their marginal cost of using the machines fell below the marginal cost of acquiring and using the new technolo-
gy. we know from the jury's answers to the interrogatories that as late as 1964, eight years after the agreement, that the jury found no distinct market in the sale of new convenience office copiers, presumably because machines embodying non-plain-paper technology continued to be sold in substantial numbers. The jury did find such a "relevant market" in 1969. Sometime after 1964, the competitive technologies ceased to be important in the marketplace for new equipment. At that point in time, Xerox may have faced a situation something like this:

![Figure 4](image-url)
This diagram shows a demand curve with a demand curve that slopes downward, turns level, and then slopes down again. The first region represents the demand by those customers who can sue only plain-paper copiers. The second region represents customers who have a choice between Xerox and continuing to use the old technology. And the final region represents demand by customers who have no use for copiers at a price equal to the marginal cost of using the old technology, but who would sue copiers if the price were lower. The way the marginal cost curve is drawn, Xerox will never be able to serve these customers unless it can lower its own marginal costs.

Figure 4 is drawn so that Xerox's marginal cost curve forces it to produce in the region with a competitive demand curve. Why might this be? One answer, not reflected in the curve as drawn, would be that economies of scale in the manufacture and distribution of the copying machines forced Xerox's production outward. But another important factor in this situation is likely to be the inevitable obsolescence of Xerox's patents. Not only will their terms run, but there is the risk that a competitive technology may appear, making them obsolete. A marginal cost of exploiting the patent slowly is the added risk that the patent will become valueless before the market has been entered. In the case of Xerox, IBM in fact introduced the first important competitive plain-paper copier in 1970. Thus if Xerox did not expand its output to the point where it was in a position to serve the market rapidly enough, it might forever lose the opportunity to
benefit from its patent position.

This is a conjectural description of the market faced by Xerox in the late 1960's. But it has considerable relevance to the question of whether in 1956 there was a significant probability that Xerox's patent position would enable it to engage in monopoly pricing. In 1956, the combined competitive threat from older technologies and potential alternative technologies was sufficiently great that it was not very likely that Xerox would ever be able to price to a non-competitive market. If so, traditional methods of section 7 analysis would lead the courts quite easily to the answer that Xerox had no antitrust liability.

CONCLUSION

The discussion of this paper is an exploration of the question: what is the difference between a rent and a return to a monopoly. That is a basic distinction that one might expect to be well developed and articulated in the literature. As a non-economist looking to the literature for assistance, the distinction appears to be repeatedly muddled in practice. The basic distinction adopted in this paper is between a right which confers a comparative advantage in production which is sold into a market where the demand curve facing the firm has a slope of zero, and a right which confers the advantage of being able to sell into a market where the demand curve facing the firm has a negative slope. Both kinds of rights have value. Patents are most always examples of the first rather than the second class.
FOOTNOTES

1. Member of the Center for Advanced Studies, University of Virginia and Professor of Law, University of Virginia.
3. Id. at 447, n. 30.
7. My Webster's New World Dictionary of the American Language gives as meaning '3: "exclusive possession or control of something."
9. Id. at 26.
10. Id. at 44
11. Ward S. Bowman, Jr., Patent and Antitrust Law: A Legal and

12. 20 J. of Law & Econ. at 270–71.


18. I stress the section 7 argument, because that was its strongest form, but the same argument was also made under section 2 of the Sherman Act. SCM also argued that Xerox, as a monopolist, had an obligation under section 2 to offer licenses on reasonable terms to competitors. The courts dismissed this argument on the ground that one of the rights conferred by a patent is the right not to license. The result is consistent with the
welfare arguments in E.W. Kitch, "The Nature and Function of the Patent System," 20 J. of Law and Econ. 265 (1977). It could also have been reached under the argument made in this paper on the ground that the exercise of exclusive patent rights did not confer a monopoly.

19. In concluding that there was no relevant market in 1956, Judge Newman and the Second Circuit were relying on the jury's answer to the first interrogatory that no relevant market in convenience office copiers existed in 1964. By this, the jury meant that there was no market that could be monopolized since office copiers competed with duplicating machines such as mimeograph machines. This does not mean that there did not exist a market for copiers in both 1956 and 1964 that could be monopolized in the future as the result of an agreement entered into in 1956.


21. One effect of the entry of another competitor into the development and marketing of a plain paper copier might have been to raise the costs of both firms per machine of developing and marketing the machine. If that were so, then the agreement preventing such entry had the desirable property of making the development of the machine occur more efficiently. See E.W. Kitch, "The Nature and Function of the Patent System," 20 J. of Law and Econ. 265, 276-77.
22. If SCM had a comparative advantage in the distribution of a plain paper copier because of its existing distribution and service organization and customer contacts, then it might have been able to gain some return to those assets above the expense of the royalty if it had obtained a license. Because Xerox exploited the patent internally, SCM could argue that it was denied the opportunity to make this use of its marketing organization.
OCTOBER 1982
REVISED FEBRUARY 1983

PATENTS, LICENSING, AND ANTITRUST

CHRISTOPHER D. HALL
SIMON FRASER UNIVERSITY

The patent grant may be related to "monopoly," inventive activity, or to the methods of organization chosen to produce goods and ideas. Much has been said of the first two. This paper examines only the latter. The general argument is simply that patents serve to lower costs of renting ideas. Thus when greater decentralization of research and production is desired, patenting increases. Implications of the argument are compared to associations between antitrust actions and patent rates in the U.S. and in Canada.

Two empirical tasks are undertaken. One shows a history of Department of Justice antitrust actions involving patents, the other incorporates the former to explain changes in patent statistics. The history reveals a pattern of disproportionate representation of cases involving patent licensing with an otherwise rare variety, cross-licensing and/or patent pools, being rather commonplace in antitrust actions. Evidence is also presented showing that Justice antitrust cases involving patents are associated with reduced patenting rates, while nonpatent cases seem to encourage patenting. Changes in patenting by U.S. citizens in Canada supports the theory that "cartel" reduction may not have been a major result of the Justice actions, while sustaining the idea that patents are used to lower costs of renting innovations.

Although the present approach is at odds with interpreting Justice actions as taxing the formation of cartels, this more
traditional view is not taken lightly. Particularly when cross-licensing is involved, the cartel tax seems not improbable. Preist notes, "A cross-license agreement can be a more effective screen than a unilateral license with which to conceal an agreement to restrain trade." Therefore, a goal of this study was to assemble a numerical history of these Justice cases in order to see when licensing and cross-licensing has been an issue, and if there is a measurable correlation between such cases and patenting activity.

At the outset, my priors were that both the Justice cases and patent statistics would be too heterogeneous to meaningfully aggregate, and therefore, no consistent pattern would emerge. Upon viewing these data in conjunction with unlitigated licensing observations, a prima facie case for the proposition that antitrust actions have reduced the value of patents does seem warranted. However tenuous the data, none of it supports the thesis that Justice actions involving patents have, on balance, deterred the use of patents to disguise cartels, while these data are consistent with the transaction cost reducing function of patents.

It is not contended that these data prove nor disprove either theory regarding the use of patents. Even if this were possible, the data at hand are too suspect to perform such a task. Rather, these observations serve to question the net effect of Justice antitrust actions. Why, for example, if
 patented are used to disguise cartels, would Canadian patents granted to U.S. citizens decline when patent antitrust actions are on the rise? If disguise of cartels was an important use of patents, there should be no reduction in patenting in Canada as these activities are contested in the U.S. since the Combines Act prosecutions are not correlated with U.S. actions. On the other hand, such a decline is consistent with substitution toward greater use of secrecy in the U.S., as is implied by the alternative theory of patenting for the purpose of lowering costs of licensing.

Speaking of patents as if this defined a roughly homogeneous set of inventions, or entities of any type, is a major simplification. Even the most casual observation reveals vast differences in these rights. However, it is also observed that in some instances "the market" does act as if there is an element common to patents. In his analysis of trade secrets as property rights Cheung (1982) points to an obscure fact which lends direct support for the proposition of a common attribute of patents.

The observation exploited by Cheung is the use of submission agreements to govern possible transactions involving ideas. He says:

...firms in the United States seldom consider an unpatented idea unless the outside inventor signs a waiver form before revealing his idea. Customarily in such a form the submitter acknowledges that no confidential relationship is to be established and that he can rely for protection only on patent or copyright statutes.
Cheung's point is that patents reduce costs of selling, and renting ideas.

This point is further supported by observations of differential use of contractual arrangements when ideas protected by patent versus trade secrecy are rented (licensed). Trade secrets, for example, tend not to be rented non-exclusively, whereas a single user will generally be chosen when secrets are rented, pure patent rentals are seldom restricted to a single user. Further, secret rentals tend to use royalties more frequently than do patent agreements.\(^5\)

The lack of nonexclusive trade secret rentals suggests the difficulty in containing the unwanted spread of ideas lacking patent protection. The observation that royalty payments are more common in secrecy rentals also supports the proposition of higher transaction costs. The royalty serves several functions, including policing an owner's action which makes "close copies" of the secret available to competing users of the idea.\(^6\)

For present purposes, the most important feature of these observations is simply that behavior is predictable by knowing if patents are involved. Thus, despite great differences between patents, they seem to share a common characteristic of reducing costs of certain types of contracting.

All of these observations are related to the possible transaction of rights between firms and/or individuals. Thus, it would be a mistake to infer that patents always reduce
transaction costs. Utilizing an idea within the confines of a firm may change the relative value of patents as well as the absolute value of the status. Indeed, it may well be that patents have negative value for some uses as the patent explicitly requires a public disclosure of information but only protects (at most) specific applications. After reviewing a patent, costs of inventing unpatented close substitutes may fall.

All of this may seem obvious, but a respectable literature nevertheless exists which attempts to explain differences in patenting rates among large and small firms as the consequence of organizational disabilities of size. Within the present argument, firms are chosen or formed to protect secrecy, thus the reduced patenting rate. Smaller firms and individuals take more patents so as to reduce their costs of being small. Thus, the differential patenting rate provides a measure of the extent of integration rather than inventive success.

At the aggregate level, changes in patenting rates may be interpreted as an index of changes in organizational costs: declines in the output of patents reflecting substitution toward more integrated units producing both goods and ideas. The antitrust activities provide an opportunity to assess this proposition in that they may be classified by their "taxing" effect on the contractual forms. Cases involving patent licensing indirectly tax the value of patents by discouraging a major function of patents. Nonpatent cases tax the organization
designed to protect secrecy, encouraging the production of patents.

If changes in patent rates should not be interpreted without evaluating the influence of antitrust, the effects of antitrust actions on inventive activity, as measured by patents, should not be thought to solely measure changes in incentives to invest in research. An action involving licensing promotes the use of firms to protect secrecy. Thus, only to the extent that secrecy may lead to redundant research or other dissipating activities is there an effect on allocation.4

In the following section a summary of characteristics of Justice cases involving patents is presented. Following this history, a regression analysis of patent production using this history to explain movements in patenting is offered by which the reader may judge the empirical relevance of the transaction cost argument.

A reading of decisions related to patents and antitrust presents a picture of courts differentially discouraging the exercise of patent rights through licensing. When the form of license is a reciprocal exchange of patent rights, i.e., cross-licensing or "pooling," the courts seem even more disposed toward restraints. However, decisions vary with the particular circumstances of each case, thus a history based on these anecdotal observations is a particularly unsatisfactory base upon which to draw inferences regarding the actual state of law
or its changes over time.

Quantifying this history allows a view of the general pattern of decisions, but is, of course, subject to charges of superficiality caused by the very aggregation procedures required to overcome defects in the anecdotal approach. Patents are anything but homogeneous. The methods used to exploit these rights, be it one-way, cross, or pooling licenses, as well as own production through integration of research and production, are themselves found in such proliferation that any aggregation seems a heroic, and simplistic, undertaking. Nevertheless, if courts distinguish between these broad categories in some way that potential traders can interpret, a consistent pattern of behavior will result. Such patterns are present in the history of litigations, and these lend added support to the thesis of differential treatment.

The combination of patents by cross-license or pooling has been explicitly encouraged by the courts and federal government in some cases. The Manufacturers Aircraft Association, for example, was explicitly encouraged (in 1928) nominally for the purpose of reducing litigation and costs of building aircraft. 9 However, the bulk of activity has been applied toward reducing gains from combination of patents by license. Ellis summarizes the "position:"

Frequently one company can validly do things which several companies cannot agree to do in combination. 10

He argues that the antitrust restraints have been applied to
restrict the contractual forms by which returns may be captured. Although many of the per se condemnations do not apply, the thrust of precedent has been toward treating combinations as if no patents are involved. Indeed, it appears that combinations have almost become "per se" illegal as a consequence.

There are no statutes condemning patent combinations. As a result, we can only look at the history of cases involving combinations to judge the thesis. However, even a casual observer will be struck by the lack of direct evidence. When the government is the plaintiff in antitrust cases involving patents, they almost always win, regardless of the type of licensing, if any, at question. There are only 3 cases of some 159 cases brought since the inception of antitrust until 1977 where the defendant is acquitted! Further, even if the win percentage of the government was considerably lower, without additional information relating to the populations of combination and oneway licenses no inferences can validly be drawn.

Inference of any types requires a theory, and in this case, since we must look to indirect evidence the theory is even more important. Thus, the next task is to outline a theory relating incentives to incorporate specific provisions such as price-fixing, and tieins to the use of oneway and cross-licenses. These provisions are commonly reported as alleged violations of
antitrust, thus their association with contract types may be used to test the proposition that courts do not differentially penalize cross, or pool licensing. It will be argued, for example, that price-fixing provisions should be more valuable in cross-licenses (and pools), thus if the frequency of such allegations is higher for oneway licenses the weight of evidence lies against the nondiscrimination hypothesis.\(^{12}\)

It is most important to realize that the explicit penalties imposed by the courts may be a trivial fraction of the total cost of a Justice action. Rather, the costs of the defense itself seem to be the major penalty. The dearth of acquittals combined with the knowledge that every defense consumes resources suggests that the action itself be considered as the penalty (tax).\(^{13}\) In the final section, which considers changes in patent production over time, this will be the measure of the extent of antitrust tax used. However, before abandoning the task of unearthing finer classifications of antitrust actions, what data that is available for this study regarding acquittal probabilities and allegations is analyzed.

It will be argued that these data show a disproportionate concern with licensing, although there is no direct evidence that licensing of any type raises conviction probabilities. Indirect evidence, however, suggests that cross-licensing and pooling arrangements are less resistant to antitrust actions. Not only are these latter arrangements more common in antitrust
-10-

actions than they seem to be in unlitigated environments, the nature of the charges brought suggests a tendency to view the agreement as devoid of patent protection.

The Licensing Theory

Licensing can be understood as an extension of the firm. The terms of the agreement are designed to approximate controls which would otherwise be implemented by direction and explicit monitoring within the firm. As such, the income terms, i.e., royalty, time fee, and lump sum, should not be viewed in isolation from the contract duration, degree of exclusivity, the incorporation of trade secrets, and the numerous other commonly found provisions. These provisions are interwoven to encourage licensees to act as if they owned the ideas they rent. Price-fixing and tie-ins may also be viewed in this light. Since a firm always has the alternative to hire additional inputs to utilize the idea in own-production, the combination of provisions must result in approximately the same allocation as owner use. Thus, licensing represents a form of partial integration.

Cross-licenses and pooling arrangements are essentially the same. Both involve a mutual exchange of ideas, most always patent rights rather than secrets, but the pool is characterized by a central agent. Obviously, pools tend to be larger collections of patent rights, while cross-licenses are found
where a single patent is contributed by each party. Cross-
licenses seem usually to be bilateral whereas pools are
multilateral, but sublicensing of cross-license rights obscures
this distinction.

One way licenses, at least in the simple form, involve an
exchange of patent rights or secrets for revenue only. There is
no exchange licensing by the licensee except for the case of
"grant backs," where licensees agree to license future
improvements, if any, obtained on the original patents. These
licenses come in a variety of forms. Some have royalties, some
use time fees and/or lump sum payments, while most use a
combination. Some licenses are exclusive, others permit
numerous users of the rights.

If the pattern of innovation in a field was correctly
anticipated, all related research could be organized by a single
firm. If the research company chose not to embody the
innovations, they might then rent use of the research through
one way licensing on a nonexclusive basis.

The effect of a cross-license is to mimic the above firm.
Research is pooled while production is conducted by more than a
single organization. The ideas will be utilized nonexclusively
in production (since there are always at least two parties to a
cross-license), which is the same result obtained with one way
nonexclusive licensing. The incentives to regulate use of the
inventions is, thus, identical to nonexclusive licensing.
Price-fixing and tie-in provisions are methods of regulating production, as are royalties, time fees, and other provisions. However, there are systematic differences between incentives to utilize regulating devices between nonexclusive and exclusive licenses, and thus between cross-licenses and oneway licenses generally.

One essential difference between exclusive and nonexclusive licenses is the degree of direct coordination of output rates. Further, since all commodities are bundles of attributes, this difference extends to the coordination of information (advertising) provided to consumers. An exclusive licensee behaves differently than nonexclusive licensees with respect to both dimensions. This difference raises the value of price stipulations in exclusive contracts relative to royalty pricing.

In situations where licensees influence the output price, neither price-fixing nor royalties make sense within the classical production framework. If the license is exclusive a royalty raises marginal cost above that faced by own production. Even if there are mysterious unique factors owned only by the licensee, the licensor could purchase these then use the lower marginal cost, i.e., zero royalty, method. The maximum return collectable by a royalty must be reduced by the pricing method. Price-fixing would be superfluous, since the royalty free licensee would behave exactly as an owner user. The licensor would simply charge a lump sum fee to raise average
costs to the level of average revenue.

If the innovation is licensed nonexclusively, price-fixing may be used to reduce costly "over production." Suppose, for example that an innovation owner would utilize a dozen plants to produce without licensing. Instead of own production, however, he chooses to license each plant separately. Further suppose that the output at each plant affects sales at other plants but that transportation costs reduce consumers willingness to substitute between plants. With uncoordinated pricing, the usual monopolistic competition result may pertain. Each plant expands output beyond the level an owner or exclusive licensee chooses, and price is reduced. Consequently, the return for the innovation is also reduced.

Further, if there are search costs for customers, the decentralized setting of prices leads to additional costs, further reducing the value of the innovation. As each plant selects a different price (as conditions change) consumers will spend additional time searching for price information. Some of these expenses will be the direct ones of obtaining information, but there will be others resulting from customers switching between plants. These extra search costs could have been collected under a unified pricing system.

Thus, nonexclusive licenses are expected to be more frequently found with price-fixing stipulations. As cross-licenses are all nonexclusive licenses, and oneway licenses are
both exclusive and nonexclusive, the frequency of price-fixing is expected to be higher in patent combinations. Of course, the frequency of price-fixing would be higher still if the non-exclusive contracts could be excluded from the oneway license group.

The use of tiein provisions may be due to a rather different set of reasons. On the one hand, tieins may be used as a monitoring devise which facilitates the setting of future fees in exclusive licensing. With nonexclusive licensing, tieins may be a way of coordinating factor use.

A problem with exclusive licensing is the lack of comparison. The exclusive licensee is also the exclusive source of information regarding sales. A tiein arrangement provides information which substitutes for comparison of licensee reports if counting the tied good raises the probability of finding underreported outputs and thus royalty obligations.

The monitoring problem with nonexclusive licensees is different because of the ability to compare performances. However, the lack of coordination means there may be under production of activities with positive "external effects" such as advertising which reduces information costs of consumers. The production of attributes which standardize, or promote interchangeability, may also be viewed as "positive external economies" and tend to be under produced with nonexclusive licensing. The tiein may be used to offset such unprofitable incentives.
Suppose, for example, that the innovation is a design for a video recorder which is more valuable if the complimentary factor, video tape, is provided in the proper proportion. For a consumer the relevant quantity of video tape is that held by all licensee (assuming interchangeability). Since each licensee "subsidies" all other licensees by holding inventories, each has an incentive to set this quantity at a level which is not maximizing for the licensor. As a response to such incentives, licensors may introduce a tie-in to increase inventories. Such an arrangement might be the requirement to purchase a minimum number of tapes in order to obtain a license.\textsuperscript{15}

The net balance of tie-in frequency between nonexclusive and exclusive licensing is impossible to determine without a more comprehensive theory. However, it seems clear that tie-ins are useful monitoring devices in either case. Although the exclusive licensee required somewhat more monitoring, nonexclusive licensees require more coordination. Unlike the price-fixing situation, where cross-licenses are expected to use these more often, tie-ins are expected to be approximately evenly distributed between cross and oneway licenses.

We now turn to examine the facts of antitrust prosecutions and allegations. In this section it will be shown that there seems to be greater use of both price-fixing and tie-in allegations by the government in oneway licensing, and thus indirect support for the thesis that combination of patents is less resistant to antitrust.
Allegations and Outcomes

For manageability, the attributes of each case are reduced to categories, and for the statistical analysis such categories are dichotomous. Some categories are naturally dichotomous, such as when there are allegations of price-fixing or not, but other categories are "made" dichotomous at the cost of suppressing some information. For example, the outcome category contains six descriptions including acquittal, nolo contendere, consent decree, dismissal, moot issue, and "other" convictions. These have been compressed into two essential outcomes; acquittal and dismissal form one with the remaining comprising the set "guilty." It is interesting, or perhaps startling, to discover that outright acquittal in these cases is so rare that an alternative form or definition of acquittal (including dismissals) was imposed by the facts. Only three acquittals are found in the 159 decision!

Another observation regarding these cases is less surprising since it is anticipated by our research on the history of patent antitrust cases. Combination of patents through licensing is over represented in the antitrust history. There are 51 cases where some type of combination by licensing seems clearly present, some 32 percent of all cases. More importantly however, this number relative to those cases where licensing is clearly present represents some 60 percent of the cases. (51 of 84 cases).
In earlier investigations of licensing behavior we have found that cross-licensing is not a common practice. It is an extremely difficult task to accurately estimate the population relative frequency of cross versus oneway licensing, and our sample may contain bias of many types, but observations from unlitigated licenses places the relative frequency of cross-licenses of any type to be no higher than twenty percent of all licenses. We have approximately 30 cross-licenses (including pools) from approximately 150 licenses involving patents, and because about two thirds of these agreements come from S.E.C. sources where only "major" events are required to be published, combinations of patents seem more likely to be over-represented in our sample. 18

If we use the 20 percent figure as a reference, the antitrust history of a minimum of 60 percent combination agreements illustrates a rather dramatic reversal. Not only is the combination of patents over represented in the antitrust actions, so too is licensing. If we consider only those cases where we feel confident to rely on the court's description of the facts or where we have other corroborating evidence such as case studies which we or others have provided, we find that 84 of the 159 cases (53%) involve some type of a rental agreement for the use of patent rights. We can only speculate as to the relative frequency of licensing for all patents, but we would be surprised to find that even a tenth of this proportion (5%) represents the actual licensing frequency of patents in general.
Such an estimate is not entirely speculation on our part. In earlier studies of transaction costs in licensing, we discovered that it is much more difficult to locate patents than patent rentals, despite the fact that the same patent(s) are rented to numerous individuals. The following table summarizes the types of ownership or license structures found in the 159 cases:

<table>
<thead>
<tr>
<th>License Type</th>
<th>Number</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>oneway</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td>cross</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>oneway and cross</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>pool</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>pool and cross</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>cross or pool</td>
<td>51</td>
<td>32</td>
</tr>
<tr>
<td>some type of license</td>
<td>84</td>
<td>53</td>
</tr>
<tr>
<td>acquisition (only)</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>license structure is unclear**</td>
<td>62</td>
<td>39</td>
</tr>
</tbody>
</table>

* Note that percentages do not sum to 100 because a case often involves more than a single license type.

** May contain cases where licensing is either not important or simply not used.

The estimation of the number of cases involving licensing
may be understated by Table 1 because of difficulties in deciphering the Red and Blue Book case descriptions. In 62 instances these descriptions were insufficient to tell what type of a license was utilized. It may be that in these cases either the license was not an issue and thus was not sufficiently important to describe. Or, there may have been no license at all. To the extent that any of these cases involved licensing in an important manner, but the reporting simply failed to detail this fact, our estimate of the number of cases involving licensing is understated. Perhaps, for example, the 84 cases involving clear licensing should be supplemented with the 62 that may have involved licensing. In this case, some 146 of 159 cases (92%) involve licensing. However, it does seem more fruitful to retain the lower estimate as this tends to exclude cases where licensing would be a subsidiary aspect of the case.

That licensing is more frequently involved in these actions than would be predicted by their apparent relative abundance in unlitigated circumstances seems clear. However, the relative frequencies of cross-licensing (and pooling) is more difficult to assess. Taking only the clear descriptions, cross-licensing represents 51 of 84 cases, or 61 percent of the cases. If none of the unclear cases involved cross-licensing in an important manner, then the fraction declines to 51 of 146 or 35 percent of the cases, but is still 75 percent higher than is found in the sample of unlitigated licenses. This is surely a lower bound of
the relative frequencies for two additional reasons. First, some of the unclear cases may contain cross-licenses while in some of the remaining cases licensing of any form may not be important. Second, in the unlitigated sample, a disporportionate expense was applied to obtaining cross-licenses. These contracts were collected for reasons which are unrelated to this study, and thus no attempt was made to obtain a sample to be used in estimating the relative frequencies of license types. For this reason, the 20 percent estimate of the unlitigated relative frequency of cross-licenses seems clearly to be an upper bound. Thus, the upper bound is compared to the lower bound, biasing the comparison toward less divergent distributions in the litigated and unlitigated environments. Nevertheless, cross-licensing still appears to be (at least) 75 percent more common in antitrust actions.

In the next few pages an effort to see if cross-licensing is also associated with higher "conviction" probabilities is described. This effort has produced little, if any, results. It appears that if there are differences in conviction probabilities, a more subtle method is required to observe them. However, knowing what is currently unknown may facilitate future research. Thus, the results of the inquiry are described. Further, when this method is applied to estimating probabilities of observing price-fixing and tiein allegations, some associations are revealed. Most importantly however, is
simply the fact that licensing is heavily represented in the Justice antitrust actions.

Do the facts of antitrust decisions support the proposition that the courts distinguish between patent rights exercised through combination of patents in cross-licenses (including "pooling") as compared to oneway licenses? The resolution to this question is complicated by the selection criteria governing the cases we have to study, by the statistical problems inherent in analyzing dochotomous variables, by the overlap of license types illustrated in Table 1, and by the lack of conviction variation.

Consider the selection "problem." For a case to be pursued to a conclusion there must be an element of uniqueness about it or chances are that it would have been avoided by anticipating the outcome and changing the behavior which brought the action to court. This factor means that each case contains some attributes which are not shared with other cases, even if it is a certain and subtle combination of attributes. To the extent that such differences exist, they add "randomness" to the observations. For example, there are various forms of pricing arrangements which might be considered as price-fixing and few if any are known with certainty in 1910 to lead to conviction of an antitrust violation. However, over time as decisions accumulate special cases of pricing behavior become consolidated under a general rule(s) leading to an eventual
anticipation of the restraint and a reduction of cases involving these behaviors. This is a complicated process and only crudely captured (if at all) in the statistical analysis. Obviously, the dichotomous description, e.g., either there is or is not price-fixing alleged, is itself an imperfect device to describe the many forms of behavior which might fall under such a description. Nevertheless, this makes the problem of ascertaining the probability of a price-fixing allegation tractable.

As may be seen in Table 1, many of the cases involve more than a single license structure. Thus, in order to isolate the effects of each type of license on the probability of a conviction, or of observing a price-fixing allegation, a multivariate technique is required. A Chi-square analysis using two by two tables has been employed to see if license types are other than proportionately distributed between the categories of guilty and acquittal. These analyses show no pattern, but are not strictly appropriate because of the overlapping license categories. When the categories overlap, the intersection of the two sets should be treated as a new category, and thus a three by two table is the minimum necessary to compute the Chi-square probability. However, the dearth of "acquittals," however defined, reduces the number of expected elements in the cells below the minimum necessary for the test when a three by two table is used. In other words, the winning percentage of
Justice is too high given the number of cases to utilize this simple technique. For the above reasons, a logit estimation is employed.

The logit derives its name from the use of logistic probability distribution, which is approximately a "t" distribution with seven degrees of freedom. The cumulative logistic is easier to manipulate than the cumulative normal or "t" distributions and thus facilitates calculating the probabilities of drawing paired observations such as acquittal and type of license in an antitrust case. Functionally, there is little difference between logistic and normal distributions. The logistic distribution has a little more area in the tails than the normal function, and is easier to use for many types of calculations.

The logistic distribution is defined as:
\[ P = \frac{1}{1 + e^{-XB}}, \]
where \( P \) is the probability of the event and \( X \) is the vector of variables influencing the probability. \( P \) is, of course, bounded between 1 and 0, whereas \( X \) may assume any value. Estimating \( P \) directly is more difficult than the corresponding odds of the event, so that for the purpose of estimation we actually use \( \ln \frac{P}{1-P} \), or the "log odds" of the event as the dependent variable. Normally the odds of an event are expressed as the inverse of the expression, e.g., if the probability of an event is .1 the odds are \( \frac{.9}{.1} \) or "9 to 1," and this form could be
used here if the odds were minimized. Using the reciprocal allows us to maximize the inverse which has intuitive appeal since the signs are consistent with maximizing probabilities directly.

By using the log odds, the probabilities are expressed as a linear function of the variables:

\[ \ln\left( \frac{P}{1-P} \right) = \ln P - \ln(1-P) = -\ln(1+e^{-XB}) - (\ln(e^{-XB}) - \ln(1+e^{-XB})) = XB \]

A search is then made to discover that set of B's which maximize the likelihood function, \( L(B) \).

Since the log-odds are a linear function of the variables, \( X \), the probabilities are not, (another desirable attribute of the logit). The probability of an event should never become 1 since there is always some error, and increases in the probability should be increasingly difficult to discover as the probability approaches 1. Thus, the change in probability of an event must be a function of how likely the event is at the point where the change in the value of the independent variable is observed. By implicitly differentiating the likelihood function with respect to the independent variable the change in probability can be ascertained and is:

\[ \frac{\partial P}{\partial X} = \frac{3}{3}(1/(1+e^{-XB})) = BP(1-P). \]
Since $P$ is always between 0 and 1, $P(1-P)$ is always positive and thus $\frac{\partial P}{\partial x}$ always has the same sign as $B$.

With this brief background let us now turn to examine the nature of the evidence by which to examine the question of antitrust's effect on the choice of license structure, and thus the organization of research.

Consider the question of the relationship between the cross-license or patent pools and the probability of conviction. Dismissals and acquittals are again aggregated to define a group. When either was observed a value of one was assigned to the case. If a guilty verdict, no contest, consent decree, or other decree adverse to the defendants was observed the outcome was labelled 0. The nature of the license involved in the dispute was coded.

Four categories are used to describe the type of license used. If a oneway license is observed this category takes the value of one, it is zero otherwise. The same system is used for each category, and in addition to these we have added the starting date of the trial (DATE). Although these categories are usually zero when one of the others is one, this is not necessarily the case. There are times when more than one assumes the value of one. Both a cross-license and a pooling arrangement may be described as being at issue in the case. Since many of these cases are extremely complex this is not surprising. However, when the ACQUITS category is one, then
the others will be zero since this is used to describe cases when only the acquisition of patent rights is involved, e.g., no licensing. Of course, from Table 1, we see that there are only ten cases out of the 159 where the category is relevant.

The inclusion of time in these estimations is a crude attempt at isolating the effect of precedent. A further attempt along these lines is evident in the difference between the two estimations presented. In the first, the entire data set is utilized. In the second, only the first 100 cases are used, (1910-1950), and just past the peak of antitrust activity. If there was a marked difference between the two estimations, we would suspect that the effects of precedent might better be described by some type of discontinuous variable, rather than our simplistic "time." As it happens there is no qualitative difference between the estimations, e.g., all of the signs are the same.

The argument that combination of patents by licensing is discouraged by antitrust implies that the coefficient on ONeway (the oneway license category) will be zero, since these should be treated no differently than patent rights exercised solely within a firm or by an individual without licensing. The magnitude relative to its standard error is quite small, with a "t" of only -0.25. The expected signs on the CROSS and POOL coefficients are expected to be negative, e.g., reducing the probability of an acquittal. However, the cross-license coefficient is positive, but also small relative to its standard
error. The pooling coefficient has the expected sign and is somewhat larger relative to its standard error than the other categories. The sign on the acquisition only category is not predicted by the argument, and the number of observations, being only ten, is really too small to worry very much about. Lastly, the sign on DATE, the starting date of the trial is negative, and the coefficient is almost three times as large as its standard error. This suggests that the government has spent proportionately more over time to attain convictions, if precedent is important. If the two estimations are compared, it further suggests that the trend has been accelerating over time. With the first 100 cases, both the size of the time coefficient and its size relative to its standard error, increase relative to the estimation using the entire 159 cases.

The logit estimation is:

Log Odds
(Acquit)=0.076-.13Acquis-.48Pool+.14Cross-.08Oneway-.05Date
(.08) (-.23) (-1.3) (.45) (-.25) (-2.76)
asymptotic "t" in ( ).

Log likelihood = -57.33, Chi-sq. = 76.6, d.f. = 86

Data set includes all 159 cases, i.e., 1910-1977.

Below is the logit estimation using the first 100 cases:

Log Odds
(Acquit)=1.37-.32Acquis-.57Pool+.21Cross-.34Oneway-.1Date
(1.26) (-.55) (-1.27) (.54) (-.69) (-3.6)
asymptotic "t" in ( ).

Log likelihood = -34.4, Chi-sq. = 47.3, d.f. = 51

Cases from 1910-1950 (approximately).
In subsequent estimations, the effect of observing a central enforcer in a pooling or cross-license, if price-fixing was alleged, if tying arrangements were alleged, were also analyzed. No systematic patterns were observed with these additional variables. The only variable which seems significantly associated with conviction probability as defined in this manner is time.

Since the sample contains so few instances of acquittal even when dismissals are added to the category, the absence of a striking association is not surprising. Due to the absence of direct evidence regarding conviction versus licensing arrangements, another course was adopted. The more circuitus route looks at the types of allegations observed, reasoning that if greater freedom in choosing contractual provisions is granted, de facto, to oneway licensors then the nature of the charges must be more specific than the more nebulous charges of monopolization. In other words, if combining patents by license is itself illegal, or less able to escape antitrust taxation than oneway licensing, Justice will tend to bring actions against oneway licensors only when a more objective criteria may be employed.

**Price Fixing and Tying Allegations**

We next investigated the evidence related to a hypothesis suggested in the earlier section; that if patent combination is
discouraged by antitrust then oneway licensing should be associated with less elusive complaints such as price-fixing or tying allegations. Of course, there are many other possible complaints or allegations which might serve the same purpose, but these two seem comparatively straightforward to identify. Further, as oneway licenses (because of the exclusive licenses) should find price-fixing or tying less valuable than (nonexclusive) cross-licenses, there is an inherent bias toward fewer such complaints of the oneway licenses. As a consequence, if there is a measurable association between oneway licensing and these allegations, than we have added confidence in the conclusion that antitrust discourages combinations of patents. Particularly, since some of the oneway licenses are surely nonexclusive.

In testing this hypothesis it was soon discovered that even the comparatively simple allegation of price-fixing is not identified without some ambiguity. We included one case of "price leadership" in with the allegations of explicit fixing, which is a trivial problem, but the major difficulty comes from the haziness of description of the allegation after the 1940's. For some reason we found it relatively simple to discern if price-fixing was alleged prior to 1950, and after this date we are hardly ever certain of the allegation.

Using the first 100 cases we estimated the log odds of a price-fixing allegation with starting date of the trial, oneway,
cross, and pool licenses, as independent variables. The result is:

Log odds price
fix allegation = 2.4 - 0.1 Starting date + 0.7 Oneway
              (2.3) (3.2) (2.0)
              
-0.1 Cross + 0.6 Pool
(0.3)      (2.0)

Asymptotic "t" in ()

Holding the license structure constant, oneway licensing does appear to increase the probability of a price-fixing allegation, cross-licenses seem to have no effect on the probability, but pools do seem to increase the probability. The relationship between pooling and the allegation is not consistent with our hypothesis, but we offer a conjecture: pools have a central agent assigned to monitor the patent rights and it is this agent which increases the probability of the charge.

It may be that a central party is more suspicious to antitrust prosecutors, or that such an agent reduces the costs of price-fixing, or some combination of the two. However, since there are descriptions of central agents in the court records we estimated the probability of the allegation simply as a function of the starting date and the record of this agent. The result is:

Log odds price
fix allegation = 2.4 - 0.1 Starting date + 0.8 Central party
              (2.4) (3.5) (2.9)

Note that pooling is not perfectly correlated with a central party because we rely on the court's description of the facts, which does not exactly correspond to our definition of pooling.
These results show a somewhat stronger association between the presence of a central party and the allegation, and is a plausible result if a clear description of a central party indicates that the court is concerned with this agent rather than the pool or combination itself.

The other allegation considered in some detail is the tying charge. The probability of being alleged to have illegally tied goods in the use of patents (using the hypothesis that oneway licensors have no more incentive than cross-licensor's to use tying provisions) is estimated below.

Again using the log odds estimation we obtained:

Log odds tying

\[
\text{allegation} = 0.9 - 0.03 \text{ Start date} + 1.2 \text{ Oneway} - 0.2 \text{ Cross} \\
(1.0) (1.8) \quad (4.0) \quad (0.5)
\]

\[+ 0.4 \text{ Pool} \]

(1.2)

Asymtotic "t' in ().

We were able to utilize the entire period (through 1977) for this estimation as there was little difficulty in deciding when the government alleged illegal tying arrangements. The results are similar to the price-fixing allegation estimation, although the starting date and pooling variables become "less significant."

A theory of why pooling seems to be associated with
price-fixing allegations and not with tying, is not advanced. A conjecture is that the central party in pooling may be (to the court) more valuable in enforcing a price-fixing agreement while there seems little point to a central agent in a tying arrangement. This, of course depends upon the type of tying arrangement contemplated by the courts, but it does seem that a price-fixing agreement requires continual monitoring to be effective, whereas a tying arrangement does not. In any event, since we are dealing only with the allegations of these activities, a theory of the state employees is required to explain the charge, and this is not attempted here.

The five percent significance level for the correlation coefficient given 29 complete cases is .37. Only the oneway association is significant at this level (and is significant beyond the one percent level).

Below is a table summarizing the simple correlations between price-fixing and tying allegations and the characteristic of the license used in the first 100 cases:

<table>
<thead>
<tr>
<th></th>
<th>Onweay</th>
<th>Cross</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Fixing Allegation</td>
<td>.20</td>
<td>-.05</td>
<td>.19</td>
</tr>
<tr>
<td>Tiein Allegation</td>
<td>.26</td>
<td>-.12</td>
<td>-.02</td>
</tr>
</tbody>
</table>

The 0.05 significance level for the correlation coefficient is + 0.24 given the 67 cases used. The remaining 33 were deleted because of incomplete information.
The coefficients are approximately as expected although these are only simple correlations, i.e., there are no controls. Without controlling time and the overlap of license types in a case, the oneway and cross categories preserve the signs of logit estimation. Only the pooling-tiein correlation changes sign in using the more sophisticated technique. Cross-licensing seems to be treated about the same as if no license were involved with respect to either allegation, and the lack of tiein and pooling association may again be due to the self-monitoring aspects of tieins.

The following table presents the simple correlations related to tieins for the entire period:

<table>
<thead>
<tr>
<th>Tiein Allegation</th>
<th>Oneway</th>
<th>Cross</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>.60</td>
<td>-.18</td>
<td>-.21</td>
<td></td>
</tr>
</tbody>
</table>

Before moving to examine the correspondence between Justice actions and patenting rates, it is important to see the time paths of the variables of interest. In Figure 1 the paths of nonpatent and patent actions by filing date are shown. Both reach their peak during the 1940's although the patent cases show a marked downturn following the peak, while the nonpatent cases do not show a similar decline through the 1960's. All of these series stop by 1970 because the total Justice cases is taken from Posner's data which end in 1969.22
In Figure 2 total patent cases are compared to the total number of cases involving a license of some type, and to the total number of cases involving a combination of patents, i.e., cross-licensing or pooling. As may be seen, the numbers and types of licensing cases seem to move together through time. Again, the 1940's are the most active periods of Justice activity.

The fact that these series move together means that isolating the individual effects of, say combination cases, is doubtful. However, since the patent and nonpatent cases diverge, it is possible to separate these two effects. Further, in Figure 3 it may be seen that the movement of U.S. patents granted, and Canadian patents granted to U.S. citizens exhibit noticeable variation from simple time trends, reducing difficulties with spurious correlation so common in time series data.

When these time trends are combined with the analysis of the allegations and the overall frequencies of license types involved in Justice actions, the picture is one of a fairly constant mix of cases. The probability of conviction seems to rise somewhat, but since this probability begins close to one, the change seems insignificant. Relative to unlitigated licensing, combination cases are common in Justice actions, as are licensing cases in general. Again, however, there seems to be no trend in the mix of license types over the period in question.
Figure 1

ANFRE  NONPAT
15   60

10  40

5   20

NONPAT = Total Justice cases from Posner minus ANFRE (—)
ANFRE = Total number of cases involving patents (---)
ANFRE = Total Justice cases involving patents (-----)
TLIC = Total number of cases involving licensing (---)
COMC = Total number of cases involving a combination of patents (-----)
USPC  USPAT
20  80

USPAT = Total U.S. patents granted (—)
USPC = Total Canadian patents granted to U.S. citizens (---)
Patent History and Antitrust Actions

The task undertaken in this part of the paper is to lay out evidence by which transaction cost reductions may be tested in licensing of the patent, and the changes in these costs attributable to antitrust actions. In the previous sections it was argued that observations gathered from unlitigated licenses and submission agreements support the thesis that patents reduce transaction costs of licensing, but that firms may be lower transaction cost devices when secrecy is desired. We now turn to examine implications of this argument for the behavior of patenting in the U.S. and in Canada.

There are, of course, numerous difficulties in interpreting patent statistics. Some of these arise from their highly aggregated nature. Further, technology and prices of patentable ideas surely have changed over any period sufficiently long to be useful in drawing inferences from time series data. Income has also changed over this period, and the relationship between technology and income, and between technology and patents is less than fully understood.

My approach to these problems is simplistic. I assume that patents are largely a creature of transaction costs, but that other things equal, a larger economy will have more patents. The latter proposition is in part a reflection of research being a form of investment, and an assumption that this type of investment is a linear (or log linear) function of adjusted
gross national product. Absent a theory of the precise nature of this relationship, an inexpensive statistical tool is utilized.

Actions of the Justice department are assumed to be exogenous. Perhaps the number and intensity of actions is determined by "political" considerations or other mysterious sources. Further, each action is treated as if expenditures on each case are not systematically related to other variables of interest.

There is some evidence for both contentions. The number of Justice cases is not highly correlated with real GNP, nor with the number of patents as such. With respect to the expenditures assumption, it is apparent from the history that neither the length of these cases, nor the frequency of cases pursued to the Supreme Court level is systematically related to time and/or GNP.²³

One may argue that the most important variable should be the cases won by Justice, but this happens to be useless for the reasons previously discussed. However, defending against an antitrust action tends to be expensive regardless of the outcome, and thus is a tax itself. Thus the number of cases brought is assumed to be an index of the size of this tax.

There may well be some disagreement regarding the exact numbers of cases involving patents, the characteristics of these cases, and even the winning record of Justice.²⁴ Errors in
measurement of the independent variables are surely present. However, since the errors associated with measuring the number of cases should be uncorrelated with errors in measuring patent output, this should bias the coefficients toward zero, i.e., against systematic associations.

Discerning whether or not a patent is involved in a case from the summary also introduces errors, and at times even the filing date is obscured by the aggregation of cases to avoid multiple counting. Still other errors are involved in discerning the type of license structure utilized, if any, involved in the exercise of the patent rights.

Further, there may or may not be a lag or even anticipation of patent granting as antitrust actions are encountered. It may take a year (or more) for the patent examination procedure to be completed. Similarly, potential patentees may know of the actions prior to the filing date. Throughout, it is assumed that potential patentees correctly anticipate the Justice actions, i.e., "rational expectations."

The central idea of the relationship of patents to transaction costs, and thus antitrust actions, is embodied in the assumption that the number of patents is a linear combination of GNP, and the number of antitrust actions by Justice. The Justice actions are divided into patent and nonpatent cases, and among patent cases between those involving a license and not. Further, the license group is divided between those involving
cross and pool arrangements (combined in a single category) and those where one innovator owns rights, i.e., oneway licenses.

Total patents is expected to rise with GNP, and decline with increases in patent cases. Further, patenting is encouraged by nonpatent cases for two reasons. One reason for the increase is that this number is a reflection of the budget constraint of Justice. As more nonpatent cases are brought, the expenditure per patent may decline. (I was unable to obtain the dollar expenditures by Justice on antitrust cases over this seventy year period).

The second reason for the expected positive relationship between nonpatent cases and patent production follows from the argument regarding the license as a substitute form of integration. If the substitutes for integration by patent license are more heavily taxed, e.g., an increase in nonpatent cases, the demand for patents to reduce costs of licensing (integration) rises.

Extending the argument to explain the granting of patents in Canada to U.S. citizens may shed some additional light regarding substitution toward secrecy. When the patent is taxed by Justice, the patenting rate will fall for two reasons. One is simply that investments in technology are taxed. The other is that secrets are protectable. Without the substitution toward secrecy, the increase in tax should induce an increase in the export of resources used to innovate and an increase in the
flow to the U.S. of those commodities embodying this technology. The U.S. companies will increase the size of facilities abroad. This effect should be observed as a rise in the Canadian patenting rate.

The argument, however, is that firms are used to protect secrets, and that secrecy is a substitute for patents. Thus, the size of firms should increase as the tax on patents rises. However, as firms substitute toward secrecy, the patenting rate of U.S. citizens in Canada should fall; revealing the secret in Canada while protecting it in the U.S. is difficult.

The specification for this test (estimation) is the same as for the U.S. patent data. I do not attempt to hold Canadian income constant because the variable considered is the number of Canadian patents granted to U.S. citizens. The export of technology in this argument is solely for the purpose of mitigating the antitrust tax. A more complete "model" would account for a number of other factors determining the export of technology through patenting. In any event, Canadian and U.S. income are highly correlated.

The time period chosen for these estimations is the result of a combination of factors. The starting date of 1900 was selected because it provides a few, ten to be exact, years prior to the first Justice cases involving patents. The final period, 1969 was chosen because of the availability of data related to total Justice cases.²⁵
Unfortunately, collinearity of the various types of licensing cases and the total patent cases means that it is not possible to meaningfully separate the importance of the two. Nevertheless, the regression separating Justice patent cases by the type of license involved (and many cases involve several types of license forms) is shown below. However, a more detailed analysis of the specification is only shown for the regressions where the cases are only distinguished by the presence of a patent.

\[
\begin{align*}
\text{USPAT} &= 34134 + 30.4 \text{GNPA} - 1282.8 \text{NOLP} - 1702 \text{LP} + 215.6 \text{NOPAT} \\
&\quad \text{\( R^2 = 0.7 \)} \quad \text{\( F=18.1 \)} \\
&\quad \text{\( (19.9) \quad (3.7) \quad (-2.2) \quad (-3.0) \quad (2.2) \quad \text{DW}=0.66 \)}
\end{align*}
\]

\[
\begin{align*}
\text{USPAT} &= 34427.1 + 29.3 \text{GNPA} - 1132.2 \text{NOLP} - 962.1 \text{LP} - 2464.1 \text{CLP} + 219.3 \text{NOPAT} \\
&\quad \text{\( R^2 = 0.7 \)} \quad \text{\( F=14.6 \)} \\
&\quad \text{\( (19.8) \quad (3.6) \quad (-1.8) \quad (-1.0) \quad (-2.5) \quad (2.2) \quad \text{DW}=0.66 \)}
\end{align*}
\]

USPAT = total U.S. patents granted during the year  
GNPA = U.S. GNP adjusted for inflation  
NOLP = No clearly described license but a patent involved in the case  
OLP = One way licensing of a patent  
CLP = Cross-licensing or pooling of patents  
NOPAT = Nonpatent Justice case  
LP = Some type of patent license  
"t" in ()

Each of the case variables represents the number of cases filed during the year, and roughly speaking, the signs are as expected. Patent production seems to be rising with real GNP, falling with additional cases involving either patents or licensing, and rising with nonpatent Justice cases. However, the magnitudes of the estimated coefficients are difficult to assess since cross-licensing seems not to be a common practice. As the measured effect of cross-license cases is the largest of any found
among the case regressors, yet the number of cross-licenses is surely much lower than oneway or simple licensing, it seems unlikely that separation of the Justice cases by the type of contract reveals very much. Further, there may (again) be substantial errors associated with interpreting the case descriptions, leading to possible downward bias in the estimations and general obfuscation of the relationships. The size of the Durbin-Watson statistic also indicates serial correlation in the errors.

As expected, unless the Justice cases are classified by the existence of a patent there seems little association between the number of cases and the patent rate. Below is an estimation of patent production using the Justice case count and real GNP as regressors:

USPATG=33343.1+40.8USGNPA-45.8JC R²=.37 F=20.9 DW=0.39
(17.4) (5.0) (-0.5)

In the following tables, regression results are shown where no attempt is made to classify cases by license type. The variable ANFRE is the number of Justice cases involving patents, regardless of the license type (if any) described. The effects of various ad hoc "corrections" for auto correlation are also included. In all cases the signs of the estimated coefficients are as expected, but since the magnitudes change markedly depending upon the ad hoc adjustment used, the possibility of spurious correlation should not be discounted entirely. Figure 5 displays plots of the OLS estimations and observed values over the periods. In general, patenting rises with GNP and nonpatent cases, while declining with patent cases as expected.
USPAT = Number of U.S. patents granted
EST(USPAT) = OLS estimated U.S. patents granted
### Table 2
**Total U.S. Patents Granted 1900-1970**

\[ USPAT = \alpha_0 + \alpha_1 USGNPA + \alpha_2 ANFRB + \alpha_3 (JC-ANFR) \]

| \( \alpha_0 \) & \( \alpha_1 \) & \( \alpha_2 \) & \( \alpha_3 \) & \( \varphi_1 \) & \( \varphi_2 \) & \( \varphi_1 \) & \( \varphi_2 \) & \( R^2 \) & \( F \) & \( DW \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 34163.8         | 29.4            | -1498.2         | 227.3           | 0               | 0               | 0               | 0               | .50             | 24.4            | 0.66            |
| (20.1)          | (3.8)           | (-4.5)          | (2.4)           |                 |                 |                 |                 |                 |                 |                 |
| 31274.3         | 41.6            | -187.2          | 63.3            | 0.82            | 0               | 0               | 0               | .11             | 3.7             | 1.75            |
| (6.1)           | (2.9)           | (-0.6)          | (1.0)           |                 |                 |                 |                 |                 |                 |                 |
| 31691.7         | 43.0            | -252.7          | 31.9            | 0.94            | -0.17           | 0               | 0               | .13             | 4.3             | 1.92            |
| (7.1)           | (3.3)           | (-0.8)          | (0.5)           |                 |                 |                 |                 |                 |                 |                 |
| 32909.3         | 43.2            | -900.4          | 26.5            | 0               | 0               | -0.75           | 0               | .40             | 16.2            | 1.35            |
| (15.4)          | (5.7)           | (-3.1)          | (0.44)          |                 |                 |                 |                 |                 |                 |                 |
| 32155.2         | 45.3            | -713.8          | 10.4            | 0               | 0               | -0.96           | -0.36           | .32             | 12.0            | 1.75            |
| (12.7)          | (5.2)           | (-2.3)          | (0.2)           |                 |                 |                 |                 |                 |                 |                 |

Error processes assumed for \( \varphi \) values:

**AUTO 1:** \( U_t = \varphi_1 U_{t-1} + \varepsilon_t \)

**AUTO 2:** \( U_t = \varphi_1 U_{t-1} + \varphi_2 U_{t-2} + \varepsilon_t \)

**"t" in ( ).**

**MAV 1:** \( U_t = \varepsilon_t - \varphi_1 \varepsilon_{t-1} \)

**MAV 2:** \( U_t = \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} \)
Table 3

Log Total U.S. Patents Granted 1900-1970

\[
\log(\text{USPAT}) = \alpha_0 + \alpha_1 \log(\text{USGPA}) + \alpha_2 \text{ANFRE} + \alpha_3 (\text{JC-ANFRE})
\]

<table>
<thead>
<tr>
<th>(\alpha_0)</th>
<th>(\alpha_1)</th>
<th>(\alpha_2)</th>
<th>(\alpha_3)</th>
<th>(\varphi_1)</th>
<th>(\varphi_2)</th>
<th>(\varphi_1)</th>
<th>(\varphi_2)</th>
<th>(R^2)</th>
<th>(F)</th>
<th>(DW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.67</td>
<td>0.18</td>
<td>-0.04</td>
<td>0.006</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.47</td>
<td>21.6</td>
<td>0.68</td>
</tr>
<tr>
<td>(37.4)</td>
<td>(3.3)</td>
<td>(-5.3)</td>
<td>(2.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.43</td>
<td>0.22</td>
<td>-0.006</td>
<td>0.002</td>
<td>0.84</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>2.3</td>
<td>1.65</td>
</tr>
<tr>
<td>(16.6)</td>
<td>(2.1)</td>
<td>(-0.8)</td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.32</td>
<td>0.25</td>
<td>-0.007</td>
<td>0.0007</td>
<td>1.04</td>
<td>-0.25</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td>3.1</td>
<td>1.94</td>
</tr>
<tr>
<td>(19.1)</td>
<td>(2.7)</td>
<td>(-1.1)</td>
<td>(0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.21</td>
<td>0.27</td>
<td>-0.03</td>
<td>0.0008</td>
<td>0</td>
<td>0</td>
<td>-0.77</td>
<td>0</td>
<td>0.37</td>
<td>14.8</td>
<td>1.3</td>
</tr>
<tr>
<td>(35.6)</td>
<td>(5.4)</td>
<td>(-3.6)</td>
<td>(0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.15</td>
<td>0.28</td>
<td>-0.02</td>
<td>0.0004</td>
<td>0</td>
<td>0</td>
<td>-1.0</td>
<td>-0.4</td>
<td>0.30</td>
<td>10.9</td>
<td>1.72</td>
</tr>
<tr>
<td>(50.8)</td>
<td>(4.9)</td>
<td>(-2.7)</td>
<td>(0.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simple forms of autocorrelation, such as those assumed in columns 5 through 8, will not bias the coefficients in OLS estimations, but do lead to biases in the estimation of standard errors. Given the fairly large sample size, it is thus surprising to observe changes in coefficient sizes displayed in the tables. Such changes cast doubt of the validity of the ad hoc adjustments used. Despite the complications, it is encouraging to notice that in no case does a coefficient sign change with the assumed error structure.²⁷

Partly because of these autocorrelation considerations clouding interpretation of the results, other implications of the argument were sought. Specifically, if patent cases promote substitution toward secrecy, then Canadian patents granted to U.S. citizens should decline with increases in these cases. Further, the Canadian data permits a separation of alternative hypotheses related to the patent-Justice case association.

One alternative hypothesis to the one explored in this paper was discussed earlier: Justice cases actually tax disguised cartels. If this was true on average, there should be no change in the patenting rate of U.S. citizens in Canada as a consequence of Justice actions since there has not been a change in incentives to obtain patents in Canada for the purpose of screening the actions of cartels.²⁸
Of course, it is always possible to imagine strategies to conceal activities of cartels which might account for a decline in Canadian patenting as Justice actions involving patents increase. Cartels might routinely apply for patents just to pretend that the U.S. patent is actually valuable as an innovation. At this juncture, there seems no method of disposing of such an argument on either a priori or factual grounds, however unlikely it may seem.

In Tables 4 and 5 the results of regressions "explaining" the number of Canadian patents granted to U.S. citizens are presented. Again, U.S. "real" GNP, the number of patent related cases, and the number of nonpatent cases during the year, are used as explanatory variables. As with the U.S. patent regressions, the signs are as expected in the sense that patenting in Canada seems to rise with GNP and nonpatent Justice actions, and falls with the patent related cases. Despite indications of autocorrelation, and sizable changes in coefficient sizes and corrected standard errors, the same pattern of substitution seems to be present as is predicted by the transaction cost argument. In no case is the expected sign different than expected, although the statistical significance of the relationships is difficult to assess.
Table 4

Canadian Patents Granted to U.S. Citizens 1900-1970

USPC = $\alpha_0 + \alpha_1 USGNPA + \alpha_2 ANFRE + \alpha_3 (JC-ANFRE)$

<table>
<thead>
<tr>
<th></th>
<th>AUTO 1 &amp; 2</th>
<th>MAV 1 &amp; 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>R²</th>
<th>F</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>$\alpha_1$</td>
<td>$\alpha_2$</td>
<td>$\alpha_3$</td>
<td>$\phi_1$</td>
<td>$\phi_2$</td>
<td>$\phi_1$</td>
<td>$\phi_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2409.0</td>
<td>18.2</td>
<td>-464.5</td>
<td>55.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.87</td>
<td>154.6</td>
<td>.84</td>
</tr>
<tr>
<td>(7.5)</td>
<td>(12.5)</td>
<td>(-7.3)</td>
<td>(3.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2247.4</td>
<td>18.4</td>
<td>-84.6</td>
<td>23.8</td>
<td>0.86</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.36</td>
<td>14.0</td>
<td>2.38</td>
</tr>
<tr>
<td>(1.8)</td>
<td>(5.7)</td>
<td>(-1.5)</td>
<td>(2.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2775.1</td>
<td>16.9</td>
<td>-65.2</td>
<td>21.3</td>
<td>0.67</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>.26</td>
<td>9.2</td>
<td>1.95</td>
</tr>
<tr>
<td>(1.7)</td>
<td>(4.7)</td>
<td>(-1.2)</td>
<td>(1.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2238.4</td>
<td>19.1</td>
<td>-353.8</td>
<td>39.4</td>
<td>0</td>
<td>0</td>
<td>-0.59</td>
<td>0</td>
<td>.81</td>
<td>96.9</td>
<td>1.60</td>
</tr>
<tr>
<td>(5.4)</td>
<td>(12.3)</td>
<td>(-5.5)</td>
<td>(2.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2114.0</td>
<td>19.8</td>
<td>-233.6</td>
<td>22.6</td>
<td>0</td>
<td>0</td>
<td>-0.67</td>
<td>-0.35</td>
<td>.74</td>
<td>68.2</td>
<td>1.69</td>
</tr>
<tr>
<td>(4.4)</td>
<td>(11.3)</td>
<td>(-3.6)</td>
<td>(1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5

Log Canadian Patents Granted to U.S. Citizens 1900-1970

\[ \ln(\text{USPC}) = \alpha_0 + \alpha_1 \ln(\text{USGNPA}) + \alpha_2 \text{ANFRE} + \alpha_3 (\text{JC-ANFRE}) \]

"t" in ( ).

<table>
<thead>
<tr>
<th>(\alpha_0)</th>
<th>(\alpha_1)</th>
<th>(\alpha_2)</th>
<th>(\alpha_3)</th>
<th>(\varphi_1)</th>
<th>(\varphi_2)</th>
<th>(R^2)</th>
<th>F</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.64</td>
<td>0.59</td>
<td>-0.06</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.82</td>
<td>103.9</td>
<td>0.94</td>
</tr>
<tr>
<td>(20.6)</td>
<td>(10.4)</td>
<td>(-7.3)</td>
<td>(2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.70</td>
<td>0.57</td>
<td>-0.01</td>
<td>0.002</td>
<td>0.85</td>
<td>0</td>
<td>0.24</td>
<td>8.2</td>
<td>2.51</td>
</tr>
<tr>
<td>(8.2)</td>
<td>(4.4)</td>
<td>(-1.3)</td>
<td>(1.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.52</td>
<td>0.22</td>
<td>-0.01</td>
<td>0.002</td>
<td>0.68</td>
<td>0.32</td>
<td>0.003</td>
<td>1.1</td>
<td>1.92</td>
</tr>
<tr>
<td>(1.3)</td>
<td>(1.2)</td>
<td>(-1.2)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.49</td>
<td>0.61</td>
<td>-0.05</td>
<td>0.005</td>
<td>0</td>
<td>0</td>
<td>-0.43</td>
<td>0</td>
<td>1.59</td>
</tr>
<tr>
<td>(17.7)</td>
<td>(9.8)</td>
<td>(-5.3)</td>
<td>(2.0)</td>
<td></td>
<td></td>
<td></td>
<td>1.74</td>
<td>67.1</td>
</tr>
<tr>
<td>5.59</td>
<td>0.63</td>
<td>-0.03</td>
<td>0.003</td>
<td>0</td>
<td>0</td>
<td>-0.53</td>
<td>-0.32</td>
<td>0.67</td>
</tr>
<tr>
<td>(15.4)</td>
<td>(9.2)</td>
<td>(-3.5)</td>
<td>(1.3)</td>
<td></td>
<td></td>
<td></td>
<td>47.2</td>
<td>1.65</td>
</tr>
</tbody>
</table>
Conclusions

This paper has brought together various pieces of information by which the thesis of the patent as a transaction cost saving device may be judged. Examination of descriptions of Justice cases involving patents indicates that licensing and particularly cross-licensing or pooling is disproportionately represented in Justice actions. The patenting data suggests that these actions have acted as a tax on patent use by reducing the value of licensing and have encouraged the use of firms to protect secrets.

Despite the highly aggregated nature of patent statistics and case summaries, observation of market transactions of ideas suggests a common attribute of patents. If patents do reduce costs of licensing, then patent statistics should, at least in part, reflect the extent of integration by license vis-à-vis firms. Further, changes in this index should reflect the taxing effects of Justice actions on both licensing and firms.

Some interpretations of the evidence support the thesis, while some others indicate only weak relationships. However, there is no evidence found to contradict the implications of the thesis. Frankly, I am surprised to discover the latter, given the aggregative nature of these data. That such a simple theory is neither rejected entirely nor swamped by the plethora of possible confounding events, suggests the possibility of a major role of patents in assisting specialization in research and production activities through licensing.
Footnotes

1. My thanks are extended to Randy Buck and Charlotte Twight for assembling much of the data used in this study.


3. See Groech and Standbury (1979, pp. 135-194) and Green (1980, ch. 7).


5. Cheung, op. cit., Hall (1982) found that secret licenses tend to have five year durations whereas patent licenses average about 13 years. Analyses of royalty-secrecy associations also shows that even when patent rights are combined with secrecy, royalties are more frequent. Nonexclusive licensing when some type of a secret is included occurs in only 14 of 73 licenses. In pure secret rentals, 2 of 14 are nonexclusive.

6. Royalties tax a licensor's selling secrets "under the table" or simply not spending resources on protection. The royalty payment automatically declines when new users sales increase outputs and thus drive down prices.

7. Schmookler (1972, pp. 36-46).


9. However, by 1976 the government had changed positions. At this time a consent degree was filed dissolving the "pool" rather than protest the antitrust action.


11. Preist, op. cit., does not agree. He states, "Courts have not distinguished the legality of cross-licenses from unilateral patent licenses." p. 356.
12. Preist, op. cit., discusses the increased incentives to utilize price-fixing in cross-licenses. Considering the efficiency gains from price-fixing: "When firms have cross-licensed complementary patents, for example, some form of price restriction is essential for the firm to take advantage of licensing efficiencies and still gain the full monopoly return for their inventions." p. 357. The arguments presented later regarding the inherent bias in observing price-fixing may be considered additional to Preist's discussion.

13. Posner (1970, pp. 388-395) details the history of penalties for all Justice antitrust actions. The major penalty in recent years may be the private cases following a Justice action.

14. DeVany (1975) offers a model where positive external effects do not cause under production. In his cases, there is no possibility to adjust the quantity of the benefit produced for others. The present argument requires that licensees be able to free ride upon the inventory of other producers.

15. Kenney and Klein (1982) show that tieins, in their case, block booking, may be a way to conserve search expenditures of customers. This argument suggests that nonexclusive licensees will find tieins to be more valuable.

16. The case descriptions are taken from the Red and Blue books. Consent decrees are the most common outcome (99), nolo contendere (25), other conviction (25), dismissal or nolle prosequi (18), and acquittal (3). In 25 cases more than one category applies. There is only one case where the issue became moot.

17. Twilight (1980).

18. The S.E.C. reporting rule requires corporations to report events which affect 10 percent or more of the corporate assets.

19. Posner, op. cit., reports finding 165 cases involving patents or copyright between 1910 and 1969, p. 398. There are ten patent cases between 1970 and 1977 in the present sample. The difference of 16 cases may be due to the copyright cases or to a different consolidation of cases, or a combination.
20. Gould (1973), and Klein, Murphy, and Preist (1981) have analyzed the selection problem in some detail. However, these authors deal with private litigation. With the government as plaintiff, the problem seems more complicated because of the lack of a well specified objective function.

21. When all categories are zero, the background of unclear and no contract cases becomes the implicit category.


24. Posner, op. cit., pp. 381-384, shows the winning percentage to be somewhat lower for all Justice cases. Approximately 80% of the cases are won at the lower level.

25. Posner also reports estimates of F.T.C. and private antitrust actions. I could find no measurable association between these cases and patent output. It is possible that private cases following a Justice case provide the actual penalty. No attempt has been made to isolate these possible effects, nor has an attempt been made to analyze possible effects of changes in the courts willingness to overturn patents.

26. The world wars were added as dummy variables, and WWI appeared as insignificant while WWII showed some slight decline in patenting. Since I offer no theory regarding war and patents, these are not reported.

27. Rao and Miller (1971, pp. 67-77) show the effects of using ad hoc adjustments for first-order autocorrelation, and conclude that the precision of estimation is not necessarily improved via GLS estimations when, as always, RHO is estimated. Indeed, it may be better to adopt the less conventional NBER phase cycle method rather than impose the error structures used in the following tables. An initial trial using this method found in Friedman and Schwartz (1982) produced coefficients of the expected signs with magnitudes similar to the OLS estimates and no indication of remaining autocorrelation.

28. It is only necessary that the Combines Act investigation be uncorrelated with U.S. antitrust cases.
References


Friedman, Milton and Anna J. Schwartz, Monetary Trends in the United States and the United Kingdom, Chicago; University of Chicago Press, 1982.


COLLECTIVE ADMINISTRATION OF COPYRIGHT: 
AN ECONOMIC ANALYSIS

Douglas A. Smith
Department of Economics
Carleton University
Ottawa, Canada

March 1983
I Introduction

This paper examines the potential development and economic impact of copyright collectives. Copyright collectives are defined as organizations which are designed to collectively enforce some or all of the property rights associated with the copyright system. In Canada, the performing rights societies which operate in the music business, and which have been explicitly regulated under the terms of the Copyright Act since 1931, are currently the most important collectives.

Rapid advances in technology in a number of areas but particularly in photocopying, audio-visual recordings, and re-diffusion of television broadcasts have stimulated widespread interest in extending the concept of performing right societies to other areas. New technology has lowered the cost of reproducing protected works and has generated a situation in which copyright holders feel it is necessary and feasible to enforce rights which had not been much affected by earlier technologies.

The potential development of copyright collectives has been analyzed elsewhere, notably in the work of the Economic Council of Canada (1971), Keyes and Brunet (1977) and Magnusson and Nabhan (1981). The general conclusion of these studies has been that technology has altered the pattern of use of copyright works in such a way that the collective exercise of rights is necessary to protect the economic position of creators and to strengthen the incentive functions of the Copyright Act.

The intent of this paper is to analyze the potential impact of
copyright collectives within the framework of the property rights approach to the economics of copyright. Section II outlines the economic approach to the problems of copyright, Section III provides a general analysis of the collective exercise of property rights, and Section IV applies this analysis to the assessment of the potential economic impacts of copyright collectives. Section V provides the conclusions and some implications for copyright policy of the analysis.

II Property Rights and Copyright

The underlying rationale for copyright is based on the property rights of creators. For most commodities in economic analysis, producers are assumed able to capture or appropriate the returns from producing the commodity. It is, however, the nature of many creative or cultural outputs that in the absence of a system of copyright, creators would be unable to appropriate these returns. This problem of inappropriability means that in the absence of copyright, these outputs are under-produced relative to the optimal quantity in the case of complete appropriability. In this context, property rights are created on the grounds of economic efficiency, not as the result of presumed natural rights.

Under a system of copyright, authors and publishers are protected against duplication without payment by subsequent publishers. This institution increases appropriability and increases the returns to producers of cultural or creative works. However, the copyright system by removing the possibility of unauthorized copying puts the holder of
the copyright in a position of being the sole seller of the protected work. This has the effect of increasing the price that is charged for copyright material. As has frequently been pointed out, the optimal degree of copyright protection is determined by attempting to strike a balance between the welfare loss associated with insufficient output when appropriability is incomplete and the reduction in welfare associated with the reduced circulation of copyright works as the result of the higher price.

This analysis is somewhat more complicated in the Canadian context. Canada is a net importer of copyright material. The extent of this effect varies among the different areas of copyright but the principle remains that some portion of the increase in producer's surplus will go to non-residents. If we are using social benefit-cost analysis from the point of view of Canada, this portion of the increase in producer's surplus can no longer be counted as an offset to the reduction in consumer's surplus in determining the optimal degree of copyright protection.

In general, however, the analysis of the optimal degree of copyright protection is essentially the same as for the optimal degree of patent protection. In both cases, the costs of restricted use of protected material must be compared with the costs of having a smaller quantity of resources devoted to the production of cultural or creative works. There is an important difference, however, which makes the system of copyright less restrictive of use than the patent system. Under the system of copyright, only the form of expression of an idea is protected. This is in contrast with the patent system in which the patent blocks
related processes based on the patented principle. Therefore, under
the copyright system, protected works will, in general, have many close
competitors even though exclusive rights to each form of expression are
granted. The potential monopoly power for individual holders of copy-
right whose works must compete with each other is in most instances not
likely to be substantial.

There is a further element in the consideration of copyright issues
which does not enter the standard analysis. The argument against copy-
right protection is that it limits the use of something that is not in-
herently scarce.¹ The point of this argument is that copyright works
once produced should be distributed at the lowest possible cost. How-
ever, as Demsetz (1969) has argued, this separation between production
and consumption is not helpful. Although there is limited inherent
scarcity in already produced works, resources to produce new works are
scarce. In the absence of copyright protection, the ex ante return ex-
pected by book publishers for example would be negative. Copiers would
face only the marginal costs of reproducing the original while the first
publisher would have to incur the same variable costs plus a series of
other costs which would include payment to the author.

In the property rights approach, the function of economic analysis
is to indicate the impact, in terms of social benefits and costs, of
different allocations of property rights. Property rights allocations
affect benefits and costs because property rules have an impact on the

¹. Hindley (1971).
extent and characteristics of transactions among individuals. The existing Copyright Act defines the rights of creators but the case has been strongly made that this Act and the rights which it defines have been overtaken by technology and that alterations in property rights in a new Act can produce beneficial social results. The way that this process has worked in the past and is expected to work in the future is effectively illustrated in the following quotation from a leading reference work on the Canadian system of copyright.

"the growth of the law of copyright protection has closely followed the development of mechanical means of reproduction. Literary copyright was protected only after the invention of printing; artistic copyright was only established with the expansion and use of engraving and lithography. The rights to exclusive reproduction by means of records and other mechanical contrivances followed shortly after the invention of these devices. The Copyright Act was amended in 1931 to include within the definition of copyright the sole right to communicate a work by radio communication. No copyright legislation affecting television and rédiffusion appeared in any Commonwealth country until 1956 although these rights were accepted by the courts as falling within copyright before they received statutory recognition. Just as the English courts adapted their judgments to the new medium of broadcasting, a word that did not appear in the Imperial Act of 1911, so it is probable that new and presently unknown methods and means of reproducing works will be recognized by judicial tribunals before legislation can be framed to meet the need."

(Fox, 1967, p. 40)

III The Collective Exercise of Property Rights

A copyright collective can be defined as an organization established to enforce property rights which cannot be economically enforced on an individual basis. To anticipate the central point of this section, we would expect that collectives might fail to emerge for the following two reasons. First, there may not be a legally enforceable property
right to serve as a basis for contracts. Second, although the property right exists, the transactions costs of enforcing the right are so high that it is simply not economical to attempt to do so even on a collective basis.

Although it is possible to categorize the reasons for the absence of collectives in terms of either the absence of the underlying rights or high transactions costs, the property rights approach stresses that these issues are related. Property rights themselves are created as a response to a given environment. Changes in that environment may alter the structure of transactions costs and the set of rights which are worth establishing. The environmental factor which appears to have the greatest impact in the area of copyright is technology.

The Economic Council in its 1971 study of intellectual property argued that copyright and other intellectual property areas should be looked at within the broad framework of information policy. Many information policy problems are therefore most effectively viewed as situations in which the delineation and enforcement of property rights have failed to keep pace with information technology.

In a policy framework, the function or objective of a system of property rights is to maximize the net social benefits generated as a consequence of transactions made subject to the property rights. As a shorthand terminology, net social benefits generated by a particular


activity are often referred to as the surplus to society or simply the surplus attributable to that activity. An alternative description, therefore, of the function of a system of property rights is to minimize the dissipation of this surplus.

Consider as an illustrative example the discovery of a new oil pool. Assume that under the "best" set of property rights, the resource would generate revenues of $3 million and total costs under that definition of property rights would be $2 million. The surplus attributable to this discovery would therefore be $1 million. If on the other hand, the property rights to this resource were less clearly defined, we would expect a smaller surplus. It is well-known, for example, that the total recoverable resource is a function of the number of wells drilled. In the absence of well-defined property rights we would expect too many wells and as a result revenues would fall and costs would increase. In the absence of any enforceable rights costs and revenues would be equalized and the total surplus would be dissipated. The negative aspects of this situation can be most starkly illustrated by asserting that in this case of complete dissipation society would have been just as well off if this resource had not been discovered. The absence of property rights can transform a valuable asset into one which is worthless.  

The preceding example demonstrates that different systems of property rights produce different quantities of surplus. The best set of

4. This example abstracts from the general equilibrium effects of the disappearance of the resource in question. For a single oil field of the magnitude described in the example, this is a reasonable assumption.
property rights can be defined as that set for which the total surplus to society is maximized net of the costs of establishing and enforcing the set of property rights. For any defined set of property rights, the assumption of profit maximization is equivalent to the assumption that firms and individuals will attempt, through contractual arrangements, to minimize the dissipation of surplus subject to the costs of transacting under this given set of property rights. We will, in other words, always observe contracts designed to minimize dissipation subject to the constraint of existing property rules. This is not, however, equivalent to the surplus that would be observed under a different set of property rights. Contractual arrangements preserve the maximum possible degree of surplus under the existing rights structure but a different structure may generate even more surplus. It is a mistake to assume that evidence of market arrangements to minimize dissipation necessarily implies optimality. In a maximizing world, these contracts will capture all possible gains under the existing property rights. Many aspects of behaviour, however, may remain unaffected by contract and the economic policy problem is to determine whether a different set of property rights would permit a wider set of contracts which would further reduce the dissipation of surplus. Copyright policy in this context consists of determining which rights structure generates the greatest surplus.

IV Copyright Collectives

As the preceding section indicates, the property rights approach to economic problems revolves critically around the structure of
transactions costs. From the point of view of society, the desirability of copyright collectives therefore depends crucially on the costs of transacting in areas where collectives could potentially operate.

Transactions costs will in general define the dividing line between contracts which are negotiated and enforced on an individual basis and collective contracts. In the area of copyright, the costs of monitoring the use of copyright material appear to be of central importance and are of particular relevance in a situation of many protected works and many users.

Just as transactions costs define the boundary between private and collective contracting, they also define the boundary between areas in which it will be economical for collectives to operate and areas in which it will not be feasible to establish collectives. In addition, the methods of operation of the collective will also be affected by transactions costs. The choice, for example, between strict copyright payment and compulsory licencing will depend on the costs associated with these forms of charging users.

This approach to the emergence of copyright collectives takes as given the desirable incentive functions of the copyright system. Although payment to creators limits the circulation of existing works, longer run factors relating to the supply of creative works have led to copyright protection. If the assumptions underlying this protection are correct then, in principle, users should pay for the use of copyright material whether these uses are based on traditional patterns of use or are related to new technological developments.
At this point, however, the preceding comments about transactions costs become important. For activities in which metering is inexpensive, the link between use and returns to specific creators is strong and the copyright system retains its incentive features even when there is a movement from individual to group enforcement. However, if transactions costs lead collectives to devise payment plans which are more loosely connected to use then the incentive and efficiency justification for copyright collectives is weakened.

A further perspective on this problem is provided by the literature on the optimal pricing of public goods. This literature considers the best set of prices to charge users for a commodity with large fixed costs but low marginal cost. It is assumed that the producer is operating on the downward sloping section of the average cost curve so that setting price equal to marginal cost will not produce sufficient revenue to cover total costs.

If we assume that this output should be produced (i.e., consumers are better off with the commodity than without even if they pay more than marginal cost) then the price must be set to cover total costs. The literature on the optimal pricing of public goods indicates that the misallocation of resources generated by the deviation away from marginal cost pricing is minimized if the fixed costs are spread over the largest feasible number of users. Although there is an apparent element of fairness in spreading the burden of fixed costs, the

5. See Baumol et al. (1979).

6. Feasibility is defined in terms of transactions costs. The best divergence of prices among markets is a function of demand elasticities.
argument is entirely in terms of efficiency. The conclusion is, again, that charging new users of copyright materials through collectives appears to be a desirable extension of copyright protection provided that the costs of monitoring use and distributing revenues do not exceed revenues.

The concept of a copyright collective is subject to graphical interpretation. The central elements of the markets for intellectual property to be portrayed are the appropriability problem and the potential gains from increasing appropriability through collectives. The problem is illustrated in Exhibit 1.

In Exhibit 1, assume that we are dealing with a particular category of copyright material. Demand for this material is divided into two categories. The first category is the primary demand indicated by $D_p$ while the other category is referred to as the secondary demand. In terms of notation, the secondary demand is $D_s$ while Exhibit 1 shows the total demand $D_T$ which is the sum of $D_p$ and $D_s$. It is possible to directly calculate $D_s$ by subtracting $D_p$ from $D_T$.

In Exhibit 1, $D_p$ represents traditional uses of copyright material for which copyright holders are compensated. $D_s$, by assumption, represents "new technological uses" for which copyright holders now receive no payment. To make the example concrete, the commodity might be books, journals and other published material so that $D_s$ would represent the demand to photocopy. In the absence of appropriability problems, copyright holders would be compensated for both primary and secondary uses and the supply curve for this output would be $SS'$. However, we are assuming no payment for secondary uses so that the intersection of $SS'$
and $D_p$ determines output $q_0$ and price per unit of $p_0$. If the outputs used in the secondary market could be costlessly appropriated, we would move to the intersection of $D_T$ and $SS'$ at point $C$ and equilibrium price and quantity would rise to $p_1$ and $q_1$.  

In a property rights context, attempts to appropriate the returns from secondary uses are simply attempts to enforce property rights. The potential gains from enforcing property rights can also be analyzed in this diagram. In the absence of a collective or some other device to increase appropriability, equilibrium remains at $p_0$ and $q_0$. At this equilibrium, there is a divergence between the value to society of a further unit of output and the marginal cost of providing it. The social value of a further unit of output is $p_2$ while the marginal cost is $p_0$. This divergence exists for all units of output between $q_0$ and $q_1$.

Exhibit 1 also demonstrates the extent of the incentive for producers to organize a collective to exploit secondary demand. The shaded area $p_1p_0BC$ represents potential rent to producers that could be captured if revenues from secondary uses could be costlessly collected. The situation depicted diagrammatically illustrates the private incentives which exist to capture the returns from secondary uses.

Given the existence of these incentives, we can review possible reasons why no collectives have yet emerged outside the music business. The existence of economies of scale in monitoring and enforcement presumably means that it will not be profitable for individual producers

---

7. In this example, quantity is being measured in terms of the number of units (e.g., book titles) produced rather than in terms of use of existing titles.
to pursue secondary users. If, under the existing property rights regime, these costs are sufficiently low, we would expect to observe private companies or agencies formed to provide monitoring and enforcement services. In fact, the Harry Fox Agency whose Canadian operations were taken over in 1975 by the Canadian Musical Reproduction Rights Agency performed exactly this type of function. In other areas, however, the secondary uses of copyright material are relatively new and mechanisms to capture returns from these uses are still being considered by creators. Uncertainty about the legal status of collective enforcement appears to be a factor in leading creator groups to feel that a copyright revision is required prior to the launching of new collectives. In addition, it is clearly possible that negotiation, monitoring and enforcement costs are so high even for a collective as to exceed the potential gains from trade. That is, the dollar value of these transactions costs could exceed \( p_1 p_0 BC \).

As shown above, the extent of transactions costs is of critical importance in determining the economic viability of collectives. Possible methods of holding these costs down include establishing list prices for secondary uses and collecting according to this list rather than negotiating on a case by case basis with users. This approach has obvious attractions for secondary users and intermediating institutions such as libraries which are also concerned with minimizing the costs of access.

The replacement of prior negotiation and permission with a requirement for payment after the fact according to a price list economizes on transactions costs if there are no major enforcement problems.
in *ex post* collection. The existing performing rights societies deal with this problem through the Copyright Act and the Copyright Appeal Board which approves their tariff. Their property right is made less costly to enforce given the knowledge that the courts would routinely find in their favour for infringement if users fail to pay this tariff. This is simply a restatement of the point that legal structure and transactions costs are related.

**Constraints on the Collective Exercise of Rights**

The purpose of copyright collectives is to facilitate the exercise of copyright in situations in which the individual exercise of these rights is not economical. If there are not important barriers to this type of activity in the form of high transactions costs, the collective exercise of rights can increase the extent of appropriability and extend the incentive functions of the Copyright Act to secondary uses of copyright material. However, we are specifically concerned that the creation of new collectives should not lead to any reduction in the extent of competition among copyright holders in their primary markets. In contrast to the patent system, a frequently stressed aspect of copyright is the extent of competition among copyright holders since the property right protects only a specific form of expression of an idea. The potential benefits, in terms of increased appropriability, under a collective system of copyright enforcement may be entirely dissipated if a system of collectives facilitates monopolization of previously competitive primary markets.
Exhibit 2 can be used to illustrate this point. In Exhibit 2, \( D_T \) represents the total market demand for a particular category of copy-
right materials. \( D_T \) is the sum of \( D_P \) and \( D_S \), the primary and secondary
demand for this material. Incomplete appropriability means that \( D_T \)
exceeds \( D_P \). The result is a market equilibrium at price \( p_0 \) and quantity
\( q_0 \). This is in contrast to the ideal competitive price-quantity combina-
tion of \( p_0 \) and \( q_1 \). The welfare loss resulting from this is area ABC.

Exhibit 2 also illustrates that if a copyright collective which
operates as an effective and unregulated cartel is introduced, a price
of \( p_1 \) would be established with output of \( q_2 \). This would be in con-
trast to the ideal competitive price-quantity combination of \( p_0 \) and
\( q_1 \). The resulting welfare loss relative to the competitive ideal is
illustrated in Exhibit 2 as area ADE.

In the particular case illustrated in Exhibit 2, the welfare loss
associated with the monopoly collective (ADE) is less than the welfare
loss when returns can be appropriated only in the primary market (ABC).
In Exhibit 2, the monopoly collective has generated a welfare gain of
area BCDE. This consists of an increase in producer's surplus of FDEB
and an increase in consumer's surplus of CDF. With different initial
assumptions about the degree of appropriability and elasticity of
demand, however, the result is turned around and the monopoly distor-
tion exceeds the appropriability distortion. In addition, factor
market considerations underlying MC, which is drawn here as completely
elastic, will have an impact on the net result.

The situation as depicted in Exhibit 2 has essentially established
the copyright collective as a trade union which acts as a monopolistic
Exhibit 2
A Copyright Collective in the Primary and Secondary Market
seller of copyright works in the primary and secondary markets. However, if membership in the collective is voluntary, it is unlikely that the collective could operate as a completely effective monopolist. Individual creators would have an incentive to increase their own output at prices below $p_1$. However, the fundamental point remains and that is that the extension of collective activity from the secondary market, where it may be desirable, to the primary market will reduce the social benefits of establishing collectives. The potential for this type of action is presumably limited by competition policy.

The relevant case, therefore, is the one in which copyright holders who continue to compete in the primary market organize a collective institution to administer their rights in the secondary market. Consider as an example the case of photocopying. Authors and publishers would continue to negotiate with each other and with the users of printed material. The existing degree of competition in areas where copyright payments are now made would be unaffected. Abstracting from fair dealing and exemptions, we could now conceive of a photocopying collective which administered, on behalf of copyright holders, the rights to all printed materials that are copied. In this case again, in the absence of some form of regulation, the collective would be like a trade union which would establish a monopoly price for photocopying. The high costs of individual enforcement would in this case mean that there would likely

8. We would expect, in other words, that the degree of success of the cartel is directly related to the individual incentive to cheat on the cartel.

9. If this type of monopoly collective established a standard price or otherwise attempted to direct rents to the average member of the collective, we would expect to observe the unique or "star" performers in the market refusing to belong to the collective.
be few attempts to cheat on the cartel by way of price-cutting in individual negotiations. The welfare effect of the enhanced price in this situation would then have to be balanced against the incentive effect on authors and publishers.

The operation of an unregulated collective in the secondary market alone is analogous to the situation depicted in Exhibit 2. In this case, however, we would be dealing only with $D_s$, the secondary demand curve and output would be established at the intersection of the curve that is marginal to $D_s$ and marginal cost. The activities of an unregulated copyright collective in the secondary market would obviously fall short of the competitive ideal.

The general conclusion of this section is that a system of copyright collectives may promote economic efficiency. The efficiency gains are related to the strengthening of the incentive function in the secondary markets. However, two provisos attach to this conclusion. The first deals with transactions costs which can make collectives or any other form of enforcement uneconomic. It is extremely difficult to estimate these costs in advance - they can ultimately be judged in terms of collectives which are not established or which fail to operate on a break-even basis. The second proviso relates to the degree of competition in the markets for copyright works. Collectives which extend their operations into the primary market where individual negotiation is feasible can offset the efficiency gains achieved in the secondary market.
V Conclusions

This paper has examined the economic aspects of the operation of copyright collectives. Such collectives currently operate in the music business and there have been a variety of recommendations to extend their operation into other areas.

The dominant factor in stimulating interest in the concept of copyright collectives has been the rapid growth of new technologies which have lowered the costs of unauthorized uses of copyright works. Although in some areas, it may not be worth requiring authorization, recent developments in photocopying, home taping and cable television have all provided both opportunities and problems for copyright holders. The important aspect of these new technologies, from the point of view of copyright holders, is that they generate new or secondary uses of copyright works for which copyright holders are either not compensated or are not compensated in the same way that they are compensated for traditional uses. Collectives are designed to allow copyright holders to capture the returns from markets for their works in which individual negotiation and collection is not feasible.

The central analytical problem underlying this paper is a property rights issue. The economic analysis of property rights asserts that property rights should be structured to produce the greatest possible degree of economic efficiency. In this framework, there are no natural property rights: property rights are instead defined on a utilitarian basis. Property rights exist to structure transactions efficiently among individuals. In a property rights framework, recent changes in technology have attenuated property rights and this paper assesses the
circumstances in which an alteration in property rights or in the way property rights are enforced will produce net social benefits.

In this paper, a copyright collective has been explicitly defined in terms of economic efficiency. It has been defined as an organization to collectively enforce property rights which cannot be economically enforced individually. The point of this definition is to exclude collectives formed solely for the purpose of generating monopoly rents. For example, it is technically possible that all holders of copyrights in a particular area might agree to act collectively in dealing with all of these rights. If they were to do so, they may be able to reduce the extent of competition among rights holders and generate rents. Although copyright holders would gain from this type of arrangement, there appears to be no economic efficiency rationale for this form of organization.

In general, then, we would expect to observe collectives operating in areas where technology has generated new uses that are difficult to deal with individually. Collectives now exist in the area of performing rights for musical works but are not active elsewhere. Collectives may fail to exist for two reasons. First, there may not be a legally enforceable property right to serve as a basis for contracts. The absence of any reference in the Copyright Act to collectives outside the music business is sometimes interpreted as a limitation on the extent to which rights could legally be enforced collectively. Second, even if there is an established property right, collectives may not exist because the transactions costs of enforcing rights are too expensive relative to the gains even when collective enforcement is chosen.
The extent of transactions costs under a collective regime is identified in this paper as the central determinant of the desirability of collectives. Transactions costs determine the dividing line between individual and collective enforcement and also between collective and no enforcement. If transactions costs are sufficiently high, the efficient outcome may be that property rights are not enforced and are therefore without value. If transactions costs are low enough to make a collective feasible, they will also determine the form of operation of the collective. The choice, for example, between strict copyright payment with the collective simply acting as an agent and compulsory licensing depends on the costs associated with these alternative pricing mechanisms. It should be clear from the framework of this paper that any statements about the efficiency of collectives or of specific forms of operation must also include the costs borne by users under different institutional arrangements.

The property right conferred under the law of copyright structures transactions in a wide range of markets. Collectives could conceivably operate in many of these markets. The desirability from the point of view of economic efficiency of collectives depends crucially on transactions costs. This paper has discussed some of the costs associated with collectives in different markets but has provided no information on the extent of these costs. The difficulty of determining what the costs might be for an organization that does not yet exist is obvious. The absence of such data is, however, a serious limitation due to the crucial role of transactions costs in the analysis.
REFERENCES


Magnusson, D.N. and V. Nabhan, Exemptions under the Canadian Copyright Act, mimeo, 1981.
The Scope and Duration of the Patent Right and the Nature of Research Rivalry

D.G. McFetridge

and

M. Rafiquzzaman

Department of Economics
Carleton University

November, 1982

Paper prepared for presentation to a conference on Economic Analysis of Intellectual Property Rights to be held at the University of Western Ontario, London, Ontario, March 31, 1983. This paper has also been presented in workshops at Trent University and Concordia University. The helpful comments of participants in both workshops are gratefully acknowledged.
I. **INTRODUCTION**

A number of students of the patent system including Plant (1934), Barzel (1968), Stigler (1968), Kittl (1973) and Lourey (1979) have noted that competition for a given patent could be pushed to the point at which all privately appropriable surplus inherent in the patent is exhausted. In this case, the resource costs incurred by the contenders for the patent are just equal to the present value of the income anticipated by the patentee.

There are models, such as those of Kamien and Schwartz (1972, 1976), in which a fixed number of rivals behave in Cournot fashion and appropriable surplus is not exhausted in equilibrium. Once free entry is allowed, however, these models are also characterized by an equilibrium in which appropriable surplus is zero. In general, if inventors are homogeneous and there is free entry into inventing or, as Dales (1975) would put it, there is a common property right to invent, appropriable surplus must be exhausted ex ante.

Exhaustion of surplus may occur because research is done "too soon" (Barzel, 1968) or "too soon and too fast" (Kittl, 1973) in an effort to pre-empt rivals. Duplicate research efforts by rival inventors will also dissipate surplus. It goes without saying, however, that dissipation is limited to 100 per cent of ex ante appropriable surplus. A strategy of time-wise pre-emption which pushed the invention date forward to the point at which the present value of appropriable surplus is just equal to the opportunity cost of the resources used by the successful inventor precludes any duplicative research by rivals (Dasgupta and Stiglitz, 1980, pp. 284-287).
If all surplus is appropriable and all appropriable surplus is dissipated new inventions are not welfare-improving. Under these circumstances society will be indifferent between having and not having a given new technology. New inventions will then be welfare-improving only to the extent that part of their value is not appropriable by inventors.

There are two types of response to this conclusion. The first is that of Stigler who argues that the limitation on appropriability implied by the existing 17 year patent term implicitly recognizes the existence of surplus dissipating rivalry. Stigler takes the degree of appropriability implied by the existing patent right as optimal and solves for the amount of dissipation there must be in this case.¹

The second type of response is to assume that the term and scope of the patent right may have been and may yet be optimal when inventors are unique, it is not optimal when there is rivalry among inventors. Those taking this approach include Kamien and Schwartz (1974) and Berkowitz and Kotowitz (1979). The latter

¹For example rivalry may push research expenditures to the point that

\[ \frac{dS}{dR} = .6 \]  \hspace{1cm} (1)

where \(SA\) = present value of appropriable surplus in $ \(R\) = research costs

Suppose, however, that due to the limited length of the patent term and restrictions on discrimination, appropriable surplus represents only 60 per cent of total surplus. From society's point of view the rivalry equilibrium implies that

\[ \frac{dS}{dR} = \frac{1}{.6} \frac{dS}{dR} = .6 = 1. \]  \hspace{1cm} (2)

If the patent right rendered 60 per cent of surplus appropriable Stigler would conclude that 40 per cent of research expenditures must be duplication.
begins with the assumption that rivalry is such as to dissipate 100 per cent of appropriable surplus and then solve for the allocatively efficient degree of appropriability.

If 100 per cent of appropriable surplus is dissipated then the optimal degree of appropriability is quite low. More specifically, the optimal patent term is as short as six months.

Neither of these two approaches is particularly satisfactory. We cannot assume that the existing patent right is optimal in scope and duration any more than we can assume that all appropriable surplus is dissipated. In this paper we examine a number of arguments to the effect that all appropriable surplus is not dissipated and that the proportion of total surplus which is, in fact, appropriable has been overstated. We will then assess the extent, if any, to which the resulting reduction in the extent to which surplus is dissipated increases the optimal patent term.

II. IMPLICATIONS OF COMPETITIVE INVENTING FOR THE OPTIMAL PATENT TERM

As we indicated in Section I, with homogeneous inventors and free entry into inventing, competition for a patent will be pushed to the point at which the value of resources expended by rival inventors is just equal to the expected royalty or other income of the patentee. In this case all privately appropriable surplus resulting from an invention is dissipated in the rivalry for it.

The purpose of the patent grant is to facilitate the appropriation of the surplus resulting from a given invention by its inventor and thus to provide an incentive to invent. The longer
the patent term or the broader the scope of the patent grant the greater is the fraction of surplus which is appropriable by inventors and the greater is the value of new inventions which will be forthcoming. Thus, by transferring surplus from users to inventors the patent grant serves, up to a point, to increase total surplus.

When inventive activity is competitive the transfer of surplus from users to inventors eliminates the surplus. An increase in the fraction of surplus which is appropriable by the patentee increases the fraction of surplus dissipated in the rivalry for the patent. A longer patent term will increase the value of resources devoted to inventive activity but it will not result in the same increase in surplus as occurs when inventors are unique.

In sum, the assumption that inventors are homogeneous and that there is free entry into inventive activity serves to reduce both the economic surplus associated with a given level of inventive activity and the surplus maximizing patent term.

The effect of assuming competitive inventing on the socially optimal patent term can be illustrated diagramatically. The standard Nordhaus-Scherer optimal patent term model assumes a non-drastic process innovation which reduces unit production costs in the using industry by \( B(R) \) per cent. \( R \) is the number of units of inventive inputs employed. It is assumed that \( B'(R) > 0 \) and \( B''(R) < 0 \).

As is illustrated in Figure I, the invention will yield a royalty of \( c_0 - c_1 \lambda_0 \) per period during the term of the patent. If
FIGURE 1
Price and Output Consequences of a Cost Reducing Process Innovation when Royalties Equal the Cost Reduction
C₀ and Q₀ are set equal to one the royalty income of the patentee will be B(R) per period. At this royalty rate the patented process is just equivalent to the existing unpatented technology in terms of the cost of production it implies.

After the expiry of the patent royalties go to zero, the cost of production falls to C₁ and the output of the using industry increases to Q₁ units per period. The benefit derived by society from a process innovation which reduces production costs by B(R) per cent is given by the area of rectangle C₁C₀AB which accrues to the patentee during the life of the patent and to the customers of the using industry thereafter plus the area of triangle ABE which accrues to the customers of the using industry after the expiry of the patent.

In the case of a unique inventor an increase in the patent term increases the present value of the royalties associated with an additional percentage point of cost reduction. As a consequence the inventor's profit maximizing level of inventive activity, R, and the amount of cost reduction forthcoming, B(R), are increased. The area of rectangle C₁C₀AB which accrues to society (either the inventor or users) in perpetuating is increased as is the area of rectangle ABE which accrues to users after the expiry of the patent.

The optimal patent term is finite for two reasons. First, additional percentage points of cost reduction require successively greater amounts of research (B''(R) < 0). Second, while the longer patent term makes the area of triangle ABE greater it also postpones the date at which this surplus begins to be realized by society.
and thus reduces its present value.

In the case of competitive inventing the total value of resources expended on inventive activity is, by definition, equal to the present value of the royalties the successful inventor expects to earn. The cost of inventive activity is therefore equal to the present value of the area of rectangle $C_1C_0AB$ over the life of the patent. Surplus exists only in the amount of the area of rectangle $C_1C_0AB$ which accrues to users after the expiry of the patent.

Under competitive circumstances the additional surplus generated by an extension of the patent term will be smaller than is the case when inventors are unique. An increase in the patent term does increase the amount of cost reduction forthcoming and thus the areas of rectangle $C_1C_0AB$ and triangle $ABE$. It also postpones the date at which both are realized as surplus. The reason is that $C_1C_0AB$ is surplus only to the extent that it does not accrue to inventors. By extending the patent term one is not simply transferring the present value of $C_1C_0AB$ from users to inventors but transforming from surplus into resource cost.

Since a given extension of the patent term involves less additional surplus when inventors are competitive than when they are unique, the surplus maximizing or socially optimal patent term will be shorter in the former case than in the latter.

The magnitude of the difference between the optimal patent term with unique and competitive inventors respectively has been investigated by Berkowitz and Kotowitz (1979). The latter follow Nordhaus (1969) in finding the patent term for which social
welfare is maximized. While Nordhaus assumed that the level of inventive activity is such that marginal royalty income (in present value terms) is just equal to marginal research cost, Berkowitz and Kotowitz assume that inventive activity is carried to the point at which the present value of royalty income is equal to research cost. That is, all privately appropriable surplus is dissipated.

The welfare function which society seeks to maximize is

\[ W = \int_0^\infty B(R)e^{-\rho t}dt + \int_T^\infty \frac{1}{2}nB(R)^2e^{-\rho t}dt - sR \]  

(1)

where \( B(R) \) = area of \( C_1C_0AB \) when \( C_0 \) and \( Q_0 \) are set equal to 1'

\( \frac{1}{2}nB(R)^2 \) = area of \( ABE \) when the demand function of the using industry is linear and its elasticity in the neighbourhood of the initial equilibrium is \( n \)

\( T = \) term of the patent

\( \rho = \) private and social discount rate

Integration of (1) yields

\[ W = \frac{B(R)}{\rho} + \frac{n}{2\rho} [B(R)]^2(1-\Psi) - sR \]  

(2)

where \( \Psi = 1 - e^{-\rho T} \)

If inventors are unique each has the exclusive right to the royalties \( B(R) \) on its own cost-reducing process innovation. The activities of one inventor do not affect the royalties available to others. In this case each inventor has the profit function,

\[ \Pi = \int_0^T B(R)e^{-\rho t}dt - sR \]  

(3)

Maximization of inventor profits implies that the level of inventive activity, \( R \), will be such that
\[ B'(R) \frac{\psi}{\rho} = s. \]  

(4)

In order to determine the optimal or welfare maximizing patent term one differentiates (2) with respect to \( \psi \) subject to (4) and solves for \( \psi \). The result is

\[ \psi_M = \frac{1 + nB}{1 + nB\left(1 - \frac{1}{2}B''B'\right)} \]  

(5)

Suppose new inventors are homogeneous and there is free entry into inventing. In this case the cost of research on any given innovation is just equal to the present value of the royalties it is expected to generate. That is, \( R \) will be such that

\[ \frac{B(R)\psi}{\rho} = sR \]  

(6)

Expression (6) implies that the change in \( \psi \) results in the following change in \( R \)

\[ \frac{dR}{d\psi} = \frac{B(R)R}{\psi(B(R) - RB'(R))} \]  

(7)

The optimal patent term when inventing is competitive can be determined by differentiating (2) with respect to \( \psi \) subject to (7) and solving for \( \psi \). The result is

\[ \psi_c = \frac{1 + nB}{1 + nB\left(1 + \frac{B}{2RB'}\right)} \]  

(8)

Comparison of expressions (5) and (8) reveals that, given the concavity of \( B(R) \) which implies that \( B > RB' \), \( \psi_M \) will exceed \( \psi_c \) and thus the optimal patent term with unique inventors will exceed the optimal patent term with competitive inventors if

\[ \frac{B}{R} - B' \geq \frac{B''B}{B'} \]  

(9)
Since $B^u < 0$, whether condition (9) holds depends on the form of the $B(R)$ function. If it takes the form $B = B R^\alpha$ as Nordhaus assumed then (9) holds as an equality and $\Psi_M > \Psi_c$.

For given values of $\alpha$, $B$, $\rho$ and $\eta$, we can calculate the optimal patent term under conditions of monopoly and competitive inventing respectively. This calculation is the same as that reported by Berkowitz and Kotowitz (1979, p. 13).

The results of our optimal patent term calculations are reported in Table 1. While the optimal patent term with unique inventors is generally quite close to that which exists in Canada (17 years) the optimal patent term with competitive inventors is much shorter, often less than one year.

The question then arises as to whether, from the point of view of the world as a whole, there is anything to be gained by retaining patent of some sort. The foregoing implies from a global point of view a patent term of less than a year is optimal. As Berkowitz and Kotowitz (1982) have also shown, it is in the interest of individual countries to opt out of even this limited system.

If we are to argue that anything remotely approaching the present system, either in Canada or worldwide, can be justified on allocative efficiency grounds then the assumptions embodied in the analysis presented above must be changed. It is to this task that we now turn.
**TABLE I**

Optimal Patent Terms with Competition and Monopoly Inventing

\[ \alpha = 0.10, \ \rho = 0.10 \]

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( 0.5 )</th>
<th>( 1.0 )</th>
<th>( 1.5 )</th>
<th>( 2.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>45.0</td>
<td>38.2</td>
<td>34.3</td>
<td>31.6</td>
</tr>
<tr>
<td>0.01</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>B</td>
<td>38.2</td>
<td>31.6</td>
<td>27.8</td>
<td>25.1</td>
</tr>
<tr>
<td>0.05</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>0.10</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**\( \alpha = 0.05, \ \rho = 0.10 \)**

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( 0.5 )</th>
<th>( 1.0 )</th>
<th>( 1.5 )</th>
<th>( 2.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>37.7</td>
<td>31.0</td>
<td>27.2</td>
<td>24.5</td>
</tr>
<tr>
<td>0.01</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>31.0</td>
<td>24.5</td>
<td>21.0</td>
<td>18.5</td>
</tr>
<tr>
<td>0.05</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.10</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Lower left of cell is unique inventor, upper right of cell is competitive inventors.
III. SALVAGING AN EFFICIENCY RATIONALE FOR THE PATENT SYSTEM

If we are to argue that the present patent term and the characteristics of the patent right have some claim to optimality we must show that the appropriable surplus resulting from patented inventions is not entirely dissipated in the rivalry for the patent or that the proportion of surplus which is appropriable is less than has been assumed.

There are a number of ways in which one might argue that at least some appropriable surplus is preserved. First, one could retreat to the original Nordhaus line of argument that inventors are unique. In their study of seventeen industrial innovations Mansfield et al. (1977, p. 227) found relatively little in the way of parallel research activity. Specifically, the executives of the innovating firms generally knew of no other research organization working at the same time on roughly the same kind of innovation.

This does not imply that innovators did not act as if they feared pre-emption by some unknown rival. The failure to observe the existence of rivals may imply either that there were none or that potential rivals were pre-empted by the innovator. The finding by Mansfield et al. (p. 229) that the private rates of return on the innovations examined were relatively low is at least consistent with the argument that privately appropriable surplus was exhausted by pre-emptive invention.

In sum, an inability to list specific rival inventions cannot be taken to indicate the absence of rivalry. The uniqueness of individual inventors cannot be confirmed empirically.
Second, one could argue that the patent system has features which facilitate arrangements to constrain rivalry and prevent the complete dissipation of appropriable surplus. There are two arguments to this effect in the recent literature. The first is due to Kitch (1977) and the second to Yu (1980). These arguments will be stated and evaluated in some detail in Section IV.

Third, one could argue that while the patent system does not itself prevent or help to prevent the dissipation of surplus there are other features of the inventor's environment which do. In this paper we consider three such features. The first is the tax system. The second is the existence of post-patent competition and the third is the existence of a limited number of rivals. These environmental limitations on dissipation will be discussed and their impact on the optimal patent term examined in Sections V.1...V.4.

IV. THE PATENT SYSTEM AND ARRANGEMENTS WHICH REDUCE DISSIPATION

IV.1 The Patent as a Prospect

Kitch (1977; 1980) has argued that the patent system helps to reduce the dissipation of surplus associated with competition among rival inventors by defining an exclusive right to develop and commercialize at an early stage in the inventive process. Unproductive competition in development and commercialization is reduced if not eliminated as a consequence.

The characteristics of the patent system which enable it to fulfill this prospect function are the awarding of a patent to the first rather than the best to file, the absence of commercial viability as a patenting requirement and the definition of the patent right in terms of technological approach rather than in terms of a specific technology (1977, p. 267).
In their comment on Kitch's prospect theory McFetridge and Smith (1980) noted that, while the prospect features of the patent system may extend the domain of exclusivity, the right to invent to a patentable standard remains a common property right. Thus, although rivalry in the commercialization of a new technology might be reduced, the resulting surplus will continue to be dissipated in the rivalry for the patent itself (1980, p. 201).

McFetridge and Smith went on to argue that there is nothing in the prospect features of the patent which facilitates arrangements which constrain the rivalry for it.

The principle limitation on the extent to which the patent, with or without prospect features, can facilitate such arrangements is that the arrangements must be made before the patent comes into existence. Since the identification and definition of the technology to which a legally enforceable claim is to be attached is the outcome of the patenting process, a future patent is not simply a set of rights that will be enforceable at some future date. It is a set of rights that have yet to be defined. As such it adds nothing to the ability of innovators to specify and enforce an agreement to limit their rivalry (1980, pp. 202-3).

In his comments on the prospect thesis, Beck (1981) took a different approach. He argued that the patent grant cannot be construed as according the patentee an exclusive right to develop and commercialize. Specifically, the patentee has an exclusive right only to what he has invented, not to what he might invent (1981, pp. 5-7). While a patentee may have the right to prevent the use of improvements on his invention which have been patented by others, he does not have an exclusive right to these improvements. Indeed he may be denied access to them altogether (1981, pp. 9-10).

Beck also notes that among the many inventions where long lags between the grant of a patent and commercialization have
been observed, it is generally the case that rivalry in development and commercialization also continues to be observed (1981, pp. 16-22).

The Beck and McFetridge and Smith papers make different points about the prospect theory. Beck argues that the patent grant does not in principle and has not in practice conferred an exclusive right to develop and commercialize on the patentee. McFetridge and Smith argue that even if it did entail an exclusive right to commercialize and did so at an early stage, this would not result in the preservation of additional surplus. Beck's arguments do not preclude changes in the patent right which would endow it with prospect features. The McFetridge and Smith argument implies that such changes would be futile.

IV.2 Prior Contracting

Yu (1980, 1981) has argued that the patent system facilitates what he calls prior contracting between the potential users of an invention and its potential inventor. Yu suggests that manufacturers implicitly acting as the agents of consumers will contract with inventors to produce the inventions demanded by consumers. Competition among inventors insures that the latter are reduced to their opportunity cost (plus any rents to superior inventive skills). Competition among manufacturers ensures that they too are reduced to their opportunity cost. Surplus is not only preserved, it accrues entirely to the users of inventions.

The contract between the potential inventor and the manufacturer takes the form of a license agreement under an existing patent held by the inventor (1981, pp. 222-224). The license grants the manufacturer the right to use all relevant future inventions or improvements of that inventor. Thus, potential
inventors identify themselves by patenting either crude
versions of the type of invention required by manufacturers or
by patenting inventions in the same area of research. They
then compete for the right to license these inventions and all
future improvements or related inventions to manufacturers.

After soliciting the bids of inventors, manufacturers
become the licensees of the best inventor. The latter earns
a return equal to the cost of inventing plus a rent equal to any
advantage he might have over the next best inventor. Competition
among manufacturer-licensees ensures that the surplus resulting
from improvement and/or related inventions as they emerge
accrues to the consumers of the manufacturers' products.

This is almost too good to be true. Since there is no
restriction on the use of intramarginal inventions beyond that
which is necessary to cover the cost of inventing, surplus is
not only not dissipated it is greater than it would be if inventors
were unique.

Unfortunately, the world envisaged by Yu is too good to
be true. Consider the case on which there is a unique "best"
invention and a large number of potential inventors who differ
in quality. Suppose quality differences are manifested in the
length of time required by each inventor to produce the invention
in question. That is, the best inventor can patent the invention
first while earning economic profits. These economic profits
do not induce rival (inferior) inventors to attempt to patent
earlier because they cannot do so and cover their opportunity
cost.
In the absence of prior contracting the best inventor would be able to patent first and surplus in the amount of his differential skill rents would be preserved.

If manufacturers are able to make a prior contract with the best inventor the latter will be paid his opportunity cost which is the cost of inventing plus his differential skill rent. This rent is, however, the only appropriable surplus which preserved. The reason is that unless the inventor under contract is the first to patent the desired invention the contract is meaningless. If the superior inventor does not pre-empt them, the patent will go to one of the inferior inventors. All manufacturers will then be obliged to deal with the inferior inventor-patentee and must do so in the usual monopoly inventor context.

Thus, even in the presence of prior contracting the superior inventor contractor must patent his invention early enough to pre-empt his inferior, non-contracting rivals. This process of pre-emption will be such as to exhaust all appropriable surplus save the superior inventor's differential skill rent. Prior contracting does not prevent the type of time-wise dissipation originally described by Barzel (1968).

Yu may be assuming that the best an inferior non-contracting inventor can hope for is a bilateral monopoly situation. This would be the case if there was an exclusive right to manufacture. It is assumed, however, that manufacturing is competitive which presumably implies free entry. Thus, should he be able to patent first, the position of the non-contracting inferior inventor would be that of a simple monopolist.
Bilateral monopoly might arise in another way. The patent on the desired invention may be regarded as an improvement patent. In this case the contracting inventor may be able to deny the non-contracting inventor exclusive rights to it and vice versa. Yu's assumptions appear to preclude this outcome. Both the superior and the inferior inventors have qualified themselves for the initial round of bidding by patenting in related areas. In Yu's words,

A few patents and a crude machine may serve as credentials for an inventory (1981, p. 224).

The implication is that either, neither or both inventors may have the right to deny the other exclusive use of improvements. There is clearly no presumption that the superior, contracting inventor will have the exclusive right to improvements or can necessarily deny it to others.

In sum, observed patent license arrangements may be consistent with the existence of prior contracting. They can also be explained in other ways. More importantly, whether license arrangements can be construed as development contracts or not, they do not prevent the dissipation of all appropriable surplus in excess of differential skill rents.
V. FEATURES OF THE INVENTOR'S ENVIRONMENT WHICH REDUCE DISSIPATION

V.1 Taxes

Competition among inventors will be such as to dissipate all privately appropriable surplus, the portion of the successful inventor’s royalty income which is expected to be taxed away will therefore not be dissipated.

Our rivalry model has assumed that inventors, in aggregate, will just cover their costs _ex ante_. If all costs are deductible for tax purposes and tax liabilities are a proportion of net income then inventors in aggregate will not expect to pay any taxes and their costs will again equal their total expected royalty income.

This outcome can be illustrated in the case of the type of rivalry described by Barzel (1968). In this case an inventor pre-empts his rivals by inventing at a point in time such that earlier invention would involve economic losses. If the respective opportunity costs of the resources used by this inventor are deductible, he will pay no taxes and the existence of the tax system will not affect the time at which pre-emptive invention occurs.

An alternative rivalry model assumes that there are _n_ rival inventors who collectively make research expenditures equal to the present value of anticipated royalty income. One inventor gets the patent and can expect to pay taxes equal to some fraction of the difference between his royalty income and his own research expenditures. If the _n-1_ unsuccessful inventors are able
to write off their expenditures against other income or to sell their deductions the total value of the deductions taken will be just equal to expected royalty income and there will be no net payment of tax \textit{ex ante}.

The implication of these two examples is that if surplus is to be preserved in the form of income taxes which inventors expect to pay then either some of the costs of invention must be non-deductible for tax purposes or at least some part of the losses on unsuccessful inventive activities must be non-deductible from other income.

The existence of limitations on deductibility such as these characterize the Canadian and most other income tax systems. The opportunity cost of equity is not deductible from taxable income. As a result income must exceed costs, include the opportunity cost of equity by the amount \((\text{tr}_eE)/(1-t)\) \textit{ex ante} where \(t\) is the income tax rate, \(r_e\) is the opportunity cost of equity and \(E\) is the value of the equity outstanding.

It is also the case in Canada and in a number of other jurisdictions that research costs can be deducted only from "related income" for tax purposes (Strain, 1981). An individual inventor cannot deduct research expenses from employment income. A corporation with no taxable income cannot transfer its deduction to another corporation even an affiliated one and must therefore carry it forward with the consequent loss in its present value.

These considerations lead to the conclusion that rivalry among inventors will be carried to the point at which the cost of research is equal to anticipated royalties less the inventors'
expected tax liability. As a rough measure of the proportion of their gross income inventors might expect to pay in taxes we have taken income taxes as a fraction of personal income from the National Accounts. This proportion which we will call \( \tau \) averaged approximately 12 per cent between 1972 and 1978.

V.2 Post-Patent Competition: Theory

Much of the rivalry literature makes the implicit assumption that competition ceases once the patent is granted. In particular, the optimal patent term calculations of Berkowitz and Kotowitz (1979) assume that the successful inventor is free to extract the entire cost saving resulting from his invention in the form of royalties.

We will argue in this section that the ability of the patentee to extract royalties from the users of his invention is limited by potential competition from inferior but non-infringing substitutes. As a consequence, some of the benefits of the invention are immediately passed on to its users. The proportion of surplus which is appropriable and thus available for dissipation is reduced.

Consider the case of patented invention 1 which reduces the unit cost of production in the using industry by \( B \) per cent. There are a large number of inferior but non-infringing substitutes for 1, the invention of which becomes possible once the characteristics of 1 become known. Substitute 2 is inferior to invention 1 if there exists a royalty rate on 1, \( (1-\theta)B_1 \), \( 0 < \theta < 1 \), which makes the development of 2 unprofitable while leaving the inventor
of i with economic profits.

Any royalty in excess of \((1-\theta)B_i\) will therefore bring one or more of the potential substitutes into existence. The inventor of i cannot, therefore, expect to extract royalties in excess of \((1-\theta)\) per cent of the cost reduction which results from this invention. Cost savings in the amount of \(\theta B\) are passed on immediately to users and in present value terms the total value of resources expended in the rivalry for the patent on i cannot exceed \((1-\theta)B_i\).

Consider a second case in which the possibility of developing a non-infringing substitute for invention i becomes apparent in the course of its use. It is in the mutual interest of the inventor and the users of invention i to avoid the expenditure of resources on a duplicate invention. If the duplicate is inferior there will be some positive royalty rate \((1-\theta)B_i\) which leaves both the licensees and the patentee better off than if resources had been invested in the development of a substitute.¹

In this case it will again be known to all those competing for the patent on invention i that its appearance will make possible the invention of non-infringing substitutes and that, as a consequence, the royalty income of the successful inventor will amount to only \((1-\theta)\) per cent of the cost reduction resulting from

¹Inventing-around will be observed. Since all concerned have an incentive to avoid the development of substitute inventions which are inferior ex ante, however, the inventing around which is observed is not necessarily wasteful ex ante. Our view on this matter differs from that of Beck (1976) who takes observed inventing-around as evidence of the waste associated with the patent system.
the innovation. The balance of the cost saving, $\delta B_1$, is passed on to users and cannot be dissipated in pre-patent rivalry.

The existence of post-patent competition has the effect of increasing the surplus associated with any given level of inventive activity and, as we will demonstrate in Section V.4, increasing the optimal patent term.
V.3 Some Empirical Evidence on the Proportion of Surplus Appropriable by the Patentee

The empirical evidence we have been able to gather suggests that, insofar as cost-reducing process innovations are concerned, the patentee is likely to be able to appropriate something less than one-third of the cost reduction resulting from his invention.

Williamson (1963) surveyed the licensing practices of eleven U.S. companies and research organizations including Du Pont, General Electric and Raytheon. The evidence with respect to cost-reducing inventions was relatively meagre. Williamson found two licenses under which royalties were set at 20 per cent of the estimated net cost saving (1963, p. 36).

Williamson also examined the royalty rates awarded in thirty successful infringement suits. Of the thirty-four involved royalty rates based on the estimated savings arising from the use of the invention, the respective proportions of these savings which were awarded to the patentee in the form of royalties were 18 per cent, 100 per cent, approximately 50 per cent and over 57 per cent (1963, p. 236).

Williamson also considers the general division of rents between licensees and the licensor. He concludes:

The licensor's share of total (appropriable) surplus would generally seem to be between 20% and 50%. Two licenses with royalties based on net profits had rates of 34% and 50% respectively...This would tend to underestimate the percentage of the total surplus accruing to the licensor since "net profits" includes normal profits which are no part of the total surplus attributable to the patent rights. The same is true of the other two estimates that we came across of the normal division of profits between
the two parties. One of these is the rule of thumb mentioned by several interviewees, that a licensor could expect to obtain about 25% of the expected profit margin on a patented item...The other is the estimate in a guide for inventors that 20% and 33-40% are the approximate percentages of profits an inventor can anticipate depending on whether or not he accepts the obligation of legally defending the patent (1963, pp. 243-244).

It is important to note that if the using industry is competitive as in our model, licensees will earn no profit in excess of their normal return and their observed share of rents to the innovation will be zero. This is not because the patentee has appropriated them but because any rents left in the hands of licensees will be competed away.

Enos (1962) has studied royalty rates on catalytic cracking patents. Enos reports that the royalty rate on the Burton process was 25 per cent of the cost saving it entailed while the royalty rate on the Fluid Process was 16 per cent of the cost saving (p. 215).

Some care should be taken in interpreting Enos' findings. His cost savings are in fact the net profits resulting from the use of the process given prevailing gasoline prices. Net profits are measured after the deduction of royalty payments and depreciation charges but before the deduction of a normal return to capital.

In order to obtain a measure of royalty payments as a function of the rents resulting from the use of the process it is necessary to derive rent estimates from the profit per barrel estimates reported by Enos (pp. 306-307). Reported royalties can then be expressed as a fraction of these rents. The results of this exercise are reported in Table II.
TABLE II
Catalytic Craking Royalties as a Fraction of Economic Rents

<table>
<thead>
<tr>
<th></th>
<th>Burton Clark</th>
<th>Process Tube &amp; Tank</th>
<th>Holmes Manley</th>
<th>Houdry 1939</th>
<th>Fluid 1942 40,000BPSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit/bbl. (^1)</td>
<td>.98</td>
<td>.91</td>
<td>.19</td>
<td>.28</td>
<td>.31</td>
</tr>
<tr>
<td>+ Royalties/bbl. (^2)</td>
<td>.17</td>
<td>.10</td>
<td>.05</td>
<td>.04</td>
<td>.05</td>
</tr>
<tr>
<td>Gross Profit/bbl.</td>
<td>1.15</td>
<td>1.01</td>
<td>.24</td>
<td>.32</td>
<td>.36</td>
</tr>
<tr>
<td>- Opportunity cost of capital/bbl. (^3)</td>
<td>.06</td>
<td>.04</td>
<td>.05</td>
<td>.06</td>
<td>.07</td>
</tr>
<tr>
<td>Economic Rent/bbl.</td>
<td>1.09</td>
<td>.97</td>
<td>.19</td>
<td>.26</td>
<td>.29</td>
</tr>
<tr>
<td>Royalty + Economic Rent (%)</td>
<td>16</td>
<td>10</td>
<td>26</td>
<td>15</td>
<td>17</td>
</tr>
</tbody>
</table>

\(^1\) Enos, pp. 306-307.
\(^2\) Enos, pp. 304-305
\(^3\) (.10)(cost of cracking plant)/bbl. yr. (average). This is equivalent to Enos' depreciation charges per bbl. which are 10% straight-line per year. See pp. 304-305.
Our interpretation of Enos' data is that the royalty rates on the various catalytic cracking processes ranged from 10 to 26 per cent of the economic rents attributable to their use. These calculations are quite rough. Enos himself concedes that the lump sum royalties on the Houdry process have not been properly annualized (p. 245). They do suggest, however, that a considerable fraction of the economic rents attributable to each of these cracking processes were left in the hands of licensees.

Layton (1972, pp. 80-93) studied the licensing of the float glass process by its inventor, Pilkingtons. The float glass process reduced the overall manufacturing cost of plate glass by 25 per cent. Pilkingtons typically received a lump sum payment plus running royalties in the amount of 6 per cent of sales for the first eight years and 4 per cent of sales for the following eight years (1972, p. 88).

Using the reported 6 per cent royalty rate, 25 per cent cost saving and the 5 per cent profit rates on sales of Pilkingtons' major licensees (p. 93) we can calculate the royalty rate as a percentage of the cost saving. That is,

\[ \Delta C = 0.25C \]  
\[ S = C/(1-\Pi/S) \]  
\[ R/\Delta C = (R/S)S/\Delta C) = 0.06/(1-0.05) = 0.253 \]

Thus, running royalties for float glass were approximately 25 per cent of the cost reduction resulting from the process. Lump sum royalties could have added another five percentage points to this. Specifically, the total, undiscounted lump sum royalty receipts of Pilkingtons from 1963 to 1969 amounted to £6,950,000.
The implied annual sales of glass produced under license in 1969 is £98,050,000. The present value of annual sales at this level for ten years discounted at 10 per cent is £602,510,000. The implied percentage royalty on sales is 1.2 per cent (6,950/602,510). The implied royalty as a percentage of the cost saving is 5.1 per cent (.012/(.25)(1-.05)).

A generous estimate of Pilkingtons' royalties on the float glass process is that they amounted to 30 per cent of the cost reduction made possible by the process.

We must concede that observed royalty rates may be relatively low for reasons other than the existence of potential competition. If patents are cross-licensed the patentee may be receiving benefits from the licensee well in excess of observed royalty payments.

Alternatively, we may observe relatively low royalty rates when patent licenses are used as the basis for a cartel (Priest, 1977, pp. 326-330). There were allegations refuted by Priest (1977, pp. 364-376) that the catalytic cracking patents referred to above were the basis for a cartel arrangement among gasoline refiners. If cartel profits are attributable to a specific patent and these profits are shared with the patentee then the surplus available for dissipation in pre-patent rivalry is increased and the net surplus resulting from any given level of inventive activity is reduced.

---

1. This is the value in £ of running royalties received by Pilkingtons in 1969 divided by .06.
V.4 The Impact of Income Taxes and Post-Patent Competition on the Optimal Patent Term

The impact of post-patent competition on the amount and disposition of surplus resulting from a given rate of cost reduction is illustrated in Figure II. The price of the output of the using industry falls immediately to \( C_2 = C_0 - \theta(C_0 - C_1) \) and output expands to \( Q_2 \). As a consequence of the decline in price, additional surplus in the amount of \( (\eta\theta B^2/2) \) is realized by consumers and \( \eta\theta(1-\theta)B^2 \) is realized by the patentee during the patent term. There is, in addition, a transfer of surplus in the amount of \( \theta B \) from the patentee to consumers during the life of the patent.

This type of post-patent competition together with the taxation of inventors' incomes also has the effect of increasing the optimal patent term. The welfare function (expression (1)) becomes

\[
W = \int_0^\infty B(1+\frac{1}{2} \eta \theta^2 B + \eta \theta(1-\theta)B)e^{-\rho t}dt + \int_T^\infty \frac{1}{2} \eta(1-\theta)^2 B^2 e^{-\rho t}dt - sR
\]

\[
= \left( \frac{B}{\rho} \right) + \eta B^2 \left[ 1-\psi + \theta(2-\theta)\psi \right]/2\rho - sR
\]

(13)

where \( \psi = 1-e^{-\rho T} \)

and \( T = -\ln(1-\psi)/\rho \)

The zero after-tax profit equilibrium condition of competitive inventors is

\[
\psi(1-\theta)B(1+\eta B)/\rho = (1/(1-\tau))sR
\]

(15)

The response of competitive inventors to a change in the patent term is given by the derivative of (15) with respect to \( \psi \) which is
FIGURE II

Price and Output Effects of a Cost Reducing Process Innovation
When Royalties are Less Than the Cost Reduction
\[
\frac{dR}{d\Psi} = \frac{RB(l + \eta B)}{\Psi[B(1+\eta B) - RB'(1+2\eta B)]}
\]

The welfare maximizing patent term is determined by differentiating the welfare function (14) with respect to \( \Psi \) subject to (16), setting the result equal to zero and solving for \( \Psi \). The result is

\[
\Psi = \frac{l + \eta B}{1 + \eta B \left( \frac{1}{2} + \frac{B}{2RB'} \right) + \frac{B - RB'}{RB'} - \frac{B}{RB'}(2+\eta B)\theta - \frac{\eta B(2-\theta+\theta^2)}{2(1+\eta B)} - \frac{B}{RB'}(1-\theta)(1+\eta B)}
\]

(17)

Comparison of expressions (8) and (17) reveals that they are the same if \( \theta \) and \( \tau \) are zero. For \( 0<\theta<1 \) and \( 0<\tau<1 \), the denominator of (8) is greater than the denominator of (17) hence \( \Psi(\theta, \tau=0) \) exceeds \( \Psi(\theta, \tau=0) \) and the introduction of taxes and post-patent competition has increased the optimal patent term.

The extent to which the optimal patent term is increased depends on the functional form of B(R) and on the values assigned to the parameters \( \eta, \rho, \theta \) and \( \tau \). Following Nordhaus (1969) and Berkowitz and Kotowitz (1979) we specify B(R) as \( B R^\alpha \) and allow B to take on values .005, .01, .05 and .10 and \( \alpha \) to take on values .05 and .1. Parameters \( \theta \) and \( \tau \) are set at .75 and .12 respectively while \( \eta \) takes on values .5, 1.0, 1.5 and 2.0 and \( \rho \) takes on values of .05, .10 and .20.

The resulting optimal patent term calculations are reported in Table III. In the centre of each cell is the optimal patent term under the assumption that \( \theta=\tau=0 \). In the upper right hand corner of each cell is the optimal patent term under the assumption that
TABLE III: Optimal Patent Life

\( \alpha = 0.05, \theta = 0.75, \tau = 0.12, \rho = 10\% \)

\[ \eta \]

\[
\begin{array}{cccccc}
0.005 & .5 & 2.2 & 1.0 & 2.2 & 1.5 & 2.2 \\
   & .5 & 2.6 & .5 & 2.6 & .5 & 2.6 \\
   & 2.6 & 2.2 & 2.6 & 2.2 & 2.6 & 2.2 \\
   & .01 2.6 & .5 & 2.6 & .5 & 2.6 & .5 \\
   & .05 2.6 & 2.2 & .5 & 2.6 & .5 & 2.6 \\
   & .10 2.6 & 2.3 & .5 & 2.6 & .6 & 2.6 \\
\end{array}
\]

\[
\begin{array}{cccccc}
\alpha = 0.05, \theta = 0.75, \rho = 5\% \)

\[ \eta \]

\[
\begin{array}{cccccc}
0.005 & .5 & 4.5 & 1.0 & 4.5 & 1.5 & 4.5 \\
   & 5.1 & 1.0 & 5.1 & 1.0 & 5.1 & 1.0 \\
   & 4.5 & 5.1 & 4.5 & 5.1 & 4.5 & 5.1 \\
   & .01 4.5 & 1.0 & 4.5 & 1.0 & 4.5 & 1.0 \\
   & .05 4.5 & 1.1 & 4.5 & 1.1 & 4.5 & 1.1 \\
   & .10 4.5 & 1.1 & 4.5 & 1.1 & 4.6 & 1.1 \\
\end{array}
\]
\( \alpha = 0.1, \ \theta = 0.75, \ \tau = 0.12, \ \rho = 10\% \)

\[
\begin{array}{c|ccccc}
 & .5 & 1.0 & 1.5 & 2.0 \\
\hline
.005 & 5.1 & 5.1 & 5.1 & 5.1 \\
   & 6.0 & 6.0 & 6.0 & 6.0 \\
.01  & 5.1 & 5.1 & 5.1 & 5.1 \\
   & 6.0 & 6.0 & 6.0 & 6.0 \\
.05  & 5.1 & 5.2 & 5.2 & 5.2 \\
   & 6.1 & 6.1 & 6.1 & 6.1 \\
.10  & 5.2 & 5.2 & 5.3 & 5.3 \\
   & 6.1 & 6.1 & 6.2 & 6.2 \\
\end{array}
\]

\( \alpha = 0.05, \ \theta = 0.75, \ \tau = 0.12, \ \rho = 5\% \)

\[
\begin{array}{c|ccccc}
 & .5 & 1.0 & 1.5 & 2.0 \\
\hline
.005 & 10.2 & 10.2 & 10.2 & 10.2 \\
   & 12.0 & 12.1 & 12.1 & 12.1 \\
.01  & 10.2 & 10.2 & 10.3 & 10.3 \\
   & 12.0 & 12.1 & 12.1 & 12.1 \\
.05  & 10.3 & 10.3 & 10.4 & 10.4 \\
   & 12.1 & 12.2 & 12.2 & 12.3 \\
.10  & 10.3 & 10.4 & 10.5 & 10.6 \\
   & 12.2 & 12.3 & 12.4 & 12.5 \\
\end{array}
\]
\[ \alpha = .1, \ \theta = 0.75, \ \tau = 0.12, \ \rho = 20\% \]

\[
\begin{array}{cccccc}
\alpha & \theta & \tau & \rho & \text{Values} \\
0.05 & 3.0 & 2.6 & 3.0 & 2.6 & n \\
0.01 & 3.0 & 2.6 & 3.0 & 2.6 & n \\
0.05 & 3.0 & 2.6 & 3.1 & 2.6 & n \\
0.10 & 3.1 & 2.6 & 3.1 & 2.6 & n \\
\end{array}
\]

\[ \alpha = 0.05, \ \theta = 0.75, \ \tau = 0.12, \ \rho = 20\% \]

\[
\begin{array}{cccccc}
\alpha & \theta & \tau & \rho & \text{Values} \\
0.05 & 1.3 & 1.1 & 1.3 & 1.1 & n \\
0.01 & 1.3 & 1.1 & 1.3 & 1.1 & n \\
0.05 & 1.3 & 1.1 & 1.3 & 1.1 & n \\
0.10 & 1.3 & 1.1 & 1.3 & 1.1 & n \\
\end{array}
\]
\( \theta = .75 \) and \( \tau = 0 \). In the lower left hand corner of each cell is the optimal patent term under the assumption that \( \theta = .75 \) and \( \tau = .12 \). The optimal patent term under that assumption that \( \theta = .75 \) and \( \tau = .12 \) is generally about five times longer than when both these parameters are zero. It remains considerably shorter than the optimal patent term under the assumption that inventors are unique (see Table I).

V.5 Taxes, Post-patent Competition and International Free Riding

In a recent paper, Berkowitz and Kotowitz (1982) have demonstrated that when inventing is competitive an individual country will find it optimal to drop out of the international patent system. The welfare function for the open economy is

\[
W_i = P_i \left[ \int_0^{T_i} B e^{-\rho t} dt + \int_{T_i}^{T_w} B(1-X_i) e^{-\rho t} dt \right] \\
+ \int_{T_i}^{T_w} B X_i e^{-\rho t} dt + \int_{T_w}^{\infty} \left( B + \eta B^2 / 2 \right) C_i e^{-\rho t} dt - P_i sR
\]

(18)

where \( P_i \) = the proportion of worldwide research activity conducted in country \( i \)

= the probability that a country \( i \) inventor obtains the patent.

\( T_i \) = country \( i \)'s patent term

\( T_w \) = worldwide patent term

\( C_i \) = proportion of worldwide consumption of the output of the using industry accounted for by country \( i \)

\( X_i \) = proportion of worldwide output of the using industry accounted for by country \( i \).
In this model it is in the interest of country \( i \) to shorten its patent term, \( T_i \), relative to the worldwide patent term, \( T_w \), because, in so doing, it takes surplus out of the hands of inventors who dissipate it and puts it in the hands of local manufacturers. As a result, surplus accruing to local residents is increased.

The reason that a reduction of the local patent term increases the surplus accruing to local manufacturers is that the latter pay no royalties over the period \( T_w - T_i \) while competing against foreign manufacturers whose price reflects the fact that they must continue to pay royalties. Notice that if inventors were unique then shortening the local patent term would simply transfer surplus, in part from local inventors to local manufacturers and would not increase the welfare of country \( i \) to the same extent as it does when inventors are competitive.

Berkowitz and Kotowitz do not actually solve for the value of \( T_i \) which maximizes \( W_i \). Instead they find the values of \( C_i = X_i \) at which optimal \( T_i, T^*_i \), exceeds zero. They find that if \( T_w = 15 \) years, \( \eta = 1 \), \( \rho = 0.1 \), \( \alpha = 0.1 \) and \( B(R) = .01 \) (a 1 per cent cost reduction), the value of \( C_i \) at which \( T^*_i \) becomes positive is .875. Thus only a very large country would maintain a patent system and the "worldwide" system would collapse.

We have investigated the extent to which the introduction of taxes on inventors' royalty incomes and post-patent competition would change the Berkowitz-Kotowitz results. The calculations are long and tedious and are confined to the Appendix. Here we simply describe the model and state the conclusions we have reached.
The effect of introducing income taxes and post-patent competition on the surplus accruing to the residents of country is as a result of a B per cent cost reduction can be illustrated using Figure II provided we assume that the elasticity of worldwide demand and the elasticity of local demand are the same. A comparison of the surplus accruing to local residents in this model and in the original Berkowitz-Kotowitz model is given in Table IV.

Since in our model the price of the output of the using industry falls upon the introduction of the invention, additional surplus in the amount of \( C_1 \theta B(1+\eta B/2) \) is received by domestic consumers over \( T_w \) periods. As a consequence, the amount of surplus associated with a given rate of cost reduction or level of research is greater.

The reduction of the local patent term, \( T_1 \), relative to the world patent term, \( T_w \), does not help local consumers who pay the world price until time \( T_w \). The benefit to the local economy from the reduction in \( T_1 \) comes from the surplus which accrues to domestic manufacturers who pay no royalties but sell at a royalty inclusive world price over the period \( T_1 \ldots T_w \). In our model the gain to local manufacturers hence the incentive to reduce \( T_1 \) relative to \( T_w \) is much smaller than in the Berkowitz-Kotowitz case simply because royalties hence royalty payments avoided are a smaller fraction of the total cost reduction, \( B \).

If we assign the following parameter values: \( \theta= .75, \tau= .12, \alpha= .10, B= .01, \eta= 1, \rho= .10 \) and \( T_w = 15 \), we find that our model implies that \( T_1^* \) is greater than zero when \( P_1^* = C_1 = X_1^* = .62 \). That is,
a country with less than .62 per cent of world consumption and
production of the good produced by the industry which uses the
innovation will find it optimal to maintain no patent system.
This fraction is 25 percentage points lower than that obtained
by Berkowitz and Kotowitz but it remains higher than the share
of any country in general world output and consumption.

The minimal share required for \( T^*_1 \) to exceed zero declines
with \( T_w \). If \( T_w = 10 \) years then \( T^*_1 > 0 \) for \( P_i = C_i = X_i = .57 \). If we adopt
the results obtained in the previous section and set the world-
wide patent term at 6 years which is optimal for a closed economy
or the world as a whole under these circumstances, we would
still require \( X_i = C_i = .50 \) for \( T^*_1 > 0 \). In sum, rivalry among inventors
not only reduces the globally optimal patent term, it also provides
an incentive for even very large countries to free ride such that
international adherence to even thus limited patent regime could
not be sustained.

VI. ANOTHER LOOK AT THE NATURE OF RIVALRY FOR THE PATENT

To this point the assumption that inventing is competitive
has meant that all appropriable surplus is dissipated. Once this
assumption has been made the results are more or less predictable.
That is, despite reductions in the proportion of the surplus
which is assumed to be appropriable, the optimal patent term is
relatively short.

While most observers would concede that there is rivalry
for major patents they would have trouble with the assumption that
this rivalry is carried to the point at which appropriable surplus
TABLE IV

Comparison of Open Economy Optimal Patent Term Models

<table>
<thead>
<tr>
<th>Group</th>
<th>Model</th>
<th>Surplus Received/Period</th>
<th>No. of Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>I</td>
<td>$C_i \theta B(1+\eta B/2)$</td>
<td>$0 \ldots T_W$</td>
</tr>
<tr>
<td>Consumers</td>
<td>I</td>
<td>$C_i B(1+\eta B/2)$</td>
<td>$T_W \ldots \infty$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$0^{\dagger}$</td>
<td>$0 \ldots T_W$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$C_i B(1+\eta B/2)$</td>
<td>$T_W \ldots \infty$</td>
</tr>
<tr>
<td>Domestic Inventors</td>
<td>I</td>
<td>$p_i (1-\theta) B(1+\eta B) X_i$</td>
<td>$0 \ldots T_1$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$p_i (1-\theta) B(1+\eta B) (1-X_i)$</td>
<td>$0 \ldots T_1, T_1 \ldots T_W$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$p_i B X_i$</td>
<td>$0 \ldots T_1$</td>
</tr>
<tr>
<td>Domestic Producers</td>
<td>I</td>
<td>$(1-\theta) B(1+\eta B) X_i$</td>
<td>$T_1 \ldots T_W$</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$B X_i$</td>
<td></td>
</tr>
</tbody>
</table>

Model I assumes $0<\theta, \tau<1$

Model II assumes $\theta=\tau=0$
is exhausted. The reason is that there is likely to be some role for "inventive genius" in the research process. The ability to invent may not be unique but neither is it bestowed upon large numbers of individuals.

This implies that rivalry among inventors should be viewed within an oligopoly context rather than within the framework of perfect competition. If the ability to invent is not equally distributed over large numbers of individuals, well-endowed or superior inventors will be able to take account of the research activities of their respective rivals and "free" entry need not drive appropriable surplus to zero.

In order to explore the implications of limited rivalry on the optimal patent term we begin with a simple model in which each of n rival inventors chooses a profit maximizing level of research under the assumptions that: (a) each has a probability of 1/n of getting the patent; (b) the probability of getting the patent is not a function of research expenditures; and (c) the royalty income of the patentee depends only on the latter's own research expenditures, the research expenditures of rivals being regarded as being redundant.

In this model there are n individuals equally endowed with the ability to invent. The research activities of each inventor fully duplicate the research activities of the others. The profit function of the \( i^{th} \) inventor is then

\[
\Pi_i = \frac{\psi(B(R_i))}{n \rho} - sR_i
\]  \hspace{1cm} (19)

The \( i^{th} \) inventor chooses \( R_i \) such that
\[ \psi B' = s n \rho \]  \hfill (20)

The response of the \( i \text{th} \) inventor to a change in \( \psi \) is obtained by differentiating (20) which yields

\[ \frac{dR_i}{d\psi} = \frac{-B'}{\psi B''} \]  \hfill (21)

Since each of the \( n \) inventors chooses the same level of research activity and there is 100 per cent duplication of research activity the welfare function is

\[ W = \int_0^\infty B(R_1) e^{-\rho t} dt + \int_T^\infty n(B(R_1))^{2/2} e^{-\rho t} dt - n s R_1 \]  \hfill (22)

\[ = \frac{B}{\rho} + \frac{(1-\psi)nB^2}{2 \rho} - n s R_1 \]  \hfill (23)

The value of \( \psi \) at which (23) is maximized is obtained by differentiating (23) subject to (21), and setting the result equal to zero. The result is

\[ \psi = \frac{1 + nB}{1 + nB(1 - \frac{B''B}{2})} \]  \hfill (24)

Expression (24) is identical to (5), that is, the optimal patent term is the same as in the case of the unique inventor. The intuitive explanation for this result is that society is dealing with each of these \( n \) inventors as if they were unique. The fact that there are \( n \) inventors does not influence society's net welfare calculation at the margin.

While the number of inventors does not influence the optimal patent term it does influence the level of welfare attained at the optimal patent term. Put simply, the monopoly optimal patent
term is the best we can do regardless of n but as n increases the best we can do declines and ultimately becomes negative. The implication is that either we have a patent term which is optimal when inventors are unique or no patent at all.

VII. CONCLUSIONS

Although there have been attempts to measure the amount of rivalry for a patent (see Mansfield et al., 1977), the extent to which rivalry or fear of it results in the dissipation of the economic surplus associated with technological innovation is not known and may be unknowable.

Under these circumstances there are two possible approaches to the analysis of the adequacy of intellectual property rights. The first is to assume that the scope and duration of these rights constitutes an optimal response to whatever dissipation exists and to infer the extent of dissipation from the degree of appropriability implied by the right.

The second approach is to make some assumption regarding the degree of dissipation and then to derive the optimal degree of appropriability hence the scope and duration of the property right.

Most of the literature takes this second approach. It begins with the assumption that all appropriable surplus is dissipated and then derives a closed economy or worldwide optimal patent term which is as short as 6 months. It concludes, moreover, that no individual country would find it in its interest to maintain even this truncated patent term.
Attempts to rehabilitate the patent system involve assertions either that all appropriable surplus is not dissipated or that a smaller fraction of surplus is appropriable, given the characteristics of the patent right, than is commonly assumed.

A good case can be made for the existing patent term if one is willing to assume that the ability to invent is not equally or widely distributed. While rivalry for patents exists it need to be carried to the point at which all appropriable surplus is exhausted.

A good case can also be made for the proposition that surplus appropriable by the patentee is less than has been assumed in optimal patent term calculations. The introduction of income taxes and post-patent competition has the effect of increasing the optimal patent term by a factor of five.

We cannot make a good case for the proposition that the patent right itself has features which prevent the dissipation of appropriable surplus. Given the environment and the degree of inventor uniqueness, neither the prospect function nor the development contract function of the patent contribute to the ability of rival inventors to constrain the dissipation of surplus.

While it is correct, in our view, to assume that inventors are not unique, it is not correct to assume that inventing is characterized by large numbers and homogeneous inputs. When the optimal patent term is derived in the context of small numbers of inventors it may well approximate what now exists.
REFERENCES


AUDIO HOME RECORDING: CANADIAN COPYRIGHT IMPLICATIONS

Jim Keon
Consumer and Corporate Affairs Canada

March 1983
TABLE OF CONTENTS

I. INTRODUCTION......................................................... 1
II. LEGAL SITUATION...................................................... 2
III. ECONOMIC FUNCTION OF COPYRIGHT................................. 4
IV. IMPACT OF HOME RECORDING........................................ 6
V. SUMMARY OF HOME TAPING PRACTICES.................................. 13
VI. CANADIAN RECORDING INDUSTRY PERFORMANCE....................... 13
VII. IMPACT OF IN-HOME AUDIO RECORDING: CONCLUSIONS.............. 22
VIII. ANALYSIS OF POTENTIAL COMPENSATORY SCHEMES.................... 23
IX. CONCLUSIONS............................................................. 30
I. INTRODUCTION

The development of modern technology in the field of sound recording has enabled more and more individuals to make their own technically acceptable reproductions of copyrighted works.\(^1\) Data for Canada reveal that more than 50 percent of households now own some form of home audio recording equipment. This situation has called into question the ability of the existing copyright legislation to protect adequately the interests of copyright owners of the music and sound recordings which are duplicated. As a result, there have been appeals for amendments or revisions to copyright legislation to incorporate some form of compensation for rights owners for the use being made of their works.

This paper examines and analyzes the economic impact on copyright owners arising from the increased ownership and use of audio taping machines. Its focus is restricted to home copying which is done for private, noncommercial use. The issue of the proper balance to be reached among copyright owners and educational users who wish to have the freedom to make recordings for teaching purposes is not examined;\(^2\) nor does the paper deal with the impact of such clearly illegal practices as record or tape piracy or the illegal counterfeiting or bootlegging of material for commercial sale.\(^3\)

---

1. This paper does not deal with the issue of home video recording. The differences between home audio and video recording in terms of the technology, length of time in which the activity has been taking place, copyright interests affected by each and the manner in which each affects the market mechanisms determining the flow of copyright payments are so large as to render joint economic analysis inadequate. For such a discussion of home video recording, see Keon 1982(a).

2. This issue has been investigated as part of a comprehensive paper on copyright exemptions by Magnusson and Nabhan (1982).

3. Piracy refers to the illegal duplication and sale of music on records and tapes for commercial gain. Bootlegging is the unauthorized taping of a live performance from which copies are made and sold. Counterfeiting is the same as piracy except that the records are sold as if they were the products of the legitimate rights' owner.
The first main section of this paper contains a brief overview of the legal status of private home taping in Canada. Following this, a discussion of the economic role and purpose of copyright and the manner in which home taping may impinge upon this role is presented.

Following this, the next section develops a simple model which is used to explain the ways that home taping might be expected to affect the recording industry. Empirical data from a variety of sources are then presented and analysed in an attempt to determine the overall net impact.

The next section examines the strengths and weaknesses of various compensation payment formulas and makes recommendations regarding a preferred type of compensation scheme.

II. LEGAL SITUATION

It is not the purpose of this section to provide an exhaustive overview or discussion of the current legal status of home audio recording in Canada. The legal issue will be addressed only to the extent required to provide the framework within which the recommendation based on the paper's economic analysis must be cast.

Traditionally, the Canadian Copyright Act has provided the owner of copyright with the exclusive right to produce or reproduce his work or any substantial part thereof, subject to a number of qualifications and exceptions introduced to balance the interests of creators and users or consumers. There are no exemptions relating specifically to the broad spectrum of the reproduction of copyrighted works by individuals other than the copyright owners for personal use. Certain personal use reproductions would not, however, likely be judged to be infringing actions if they fell within the ambit of the current fair dealing provisions as set out in section 17(2) of the Act:

(17)(2) The following acts do not constitute an infringement of copyright:

(a) any fair dealing with any work for the purposes of private study, research, criticism, review or newspaper summary.

In addition, if less than the entire work is taped, this may not be judged to be substantial copying. As such, the activity would not rise to the level of infringement and the need for a fair dealing defence would not even arise. Reproductions for personal use of a substantial part of a work, or
of a work in its entirety, which do not fall within one of the
stated purposes would, therefore, appear to be infringing
actions under the present Act's provisions. 4

A recent paper by Barry Torno (1981) has concluded that
the concept of fair dealing is no longer suitable in attempt-
ting to deal with activities, such as audio and video home
taping and reprography, which involve the mass reproduction of
copyrighted works for personal use. Based on this assessment,
Torno recommends that the potential introduction of a levy on
home taping equipment should legitimize all home taping ac-
tivities. This paper accepts his conclusion. This leaves three
major responsibilities with respect to recommendations regard-
ing home taping: a) home taping could be excluded from the
category of infringing actions by means of a specific exemp-
tion; b) a system could be constructed which would provide
impunity for home tapers with remuneration to be paid to the
copyright owners by the home taping equipment manufacturers
or the individuals themselves; 5 or c) the Act could be left as
it is (i.e., home taping will continue to be an infringing
activity) with allowance made for the imposition of a levy
scheme at some future point.

The decision of whether or not to institute some form
of compensation for copyright holders as a result of home
taping is analyzed in this paper on the basis of the impact of

4. A similar view of the fair dealing provisions, as they
would apply to home taping, is presented by Nabhan
(1980). He concludes that if the decision of the Betamax
case in the United States had been made using Canadian
fair dealing criteria, the practice of home video taping
would be considered an infringement. Nabhan is of the
opinion that the fair dealing provisions would not apply
to most current audio or video home taping activity.

5. A broad injunction against the manufacturer or use of home
recording machinery would clearly be intolerable. Justice
Ferguson, in his summary of the evidence of the Betamax
case, illustrated this most succinctly:

Whatever the future percentage of legal versus
illegal home use recording might be, an
injunction which seeks to deprive the public of
the very tool of commerce capable of some
noninfringing use would be an extremely harsh
remedy, as well as one unprecedented in
copyright law. (Universal City Studios Inc.
and Walt Disney Productions v. Sony Corp. of
America).
this activity on the revenues received by affected copyright owners and the feasibility of devising and implementing an equitable compensation scheme.

III. ECONOMIC FUNCTION OF COPYRIGHT

The economic role of copyright has been well documented. A number of previous Canadian studies have explained at some length that the economic rationale for copyright protection is to increase and improve upon the rate of production of intellectual works which, in the absence of such protection, would be inadequate. Production would be inadequate since creative works in the absence of protection are easily appropriable. Any enterprising individual could obtain a copy of the copyrighted work and then proceed to manufacture and distribute it to the buying public. Since he would not have to incur the initial costs involved in creating the work, he could market it at a lower price than the original creator could. Consequently, the creator would have reduced economic incentive to produce and the output of intellectual property material would fall. Copyright protection is an attempt to rectify this market failure.

There is, however, an inherent conflict in this economic rationale: the need to trade off the improved compensation to creators (to increase their output of new works) against the concomitant diminution in the distribution of these same works because they cost more as a result of this exclusive protection. As the Economic Council of Canada stated most succinctly:

The higher returns provided to knowledge producers and processors and their innovative associates arise from higher prices to the users of the products involved (and, therefore, in smaller sales and output of them) than prevail in other circumstances. Individually, each of the new books, films and other products will be scarcer and more expensive than it would be if some more efficient and less socially costly form of incentive could be brought into play. (Economic Council of Canada, 1971, p. 36)

In Canada, the fundamental protection for copyright owners is provided in Section 3(1) of the Act, which states in part that "for the purposes of this Act", 'copyright' means

the sole right to produce or reproduce the work or any sub-
stantial part thereof in any material form whatever." Copyright owners are thus provided with the exclusive right to make and distribute their work. The optimal system for the creation and transmission of copyright protected works would be one in which, once sufficient inducement to create a work had been provided, competitive access to the mass production and distribution of the work would then be allowed. Thus, Canadian copyright legislation and indeed all intellectual and industrial property legislation is, in effect, a second best solution which attempts, through the granting of limited rights, to encourage creation of new intellectual property material without overly circumscribing the ease of access for users of this material.

The introduction of sophisticated and relatively inexpensive home reproduction equipment has greatly facilitated consumer access to and enjoyment of copyrighted works. This appears to be in line with at least one of the objectives of an effectively functioning incentive system (i.e., the production and distribution of copyright material at the lowest possible cost and in the most efficient manner). Technological improvements in home taping equipment benefit consumers, at least in the short run, by allowing greater access and increased flexibility in both the mode and format of using copyright material. This equipment has, however, limited and diminished the exclusive control over the reproduction and distribution of the work provided to the copyright owner in section 3(1) of the Act. This could potentially decrease returns to copyright owners. Thus, the challenge to the framers of copyright legislation is to ensure that such technological improvements do not undo the traditional creative incentive function of copyright. The question which needs to be answered, therefore, concerns the impact such machines have on the incentives to create or produce new works.

In discussing the economic role of copyright, it was argued above that it is crucial to determine the effect of home taping on the incentive to create copyrighted material if its impact is to be fully understood. For the recording industry, the industry most affected by audio home taping, the creative incentive aspect of copyright protection seeks to stimulate the production of master tapes. Pre-recorded tapes and records reproduce the sounds contained in the original master and are the final production output of the industry. Mass production and sale of these records and tapes produce the revenues and resulting profits, thereby providing the economic incentive to invest the time and resources in creating the master.
If the economic returns to those involved in creating sound recordings decrease for any reason, then the resources devoted to creating these recordings and the resulting number of recordings produced should likewise decline. A reduction in the rate of return will always cause a decrease in output except where a participant is said to be earning economic rents. Competitive markets cannot produce rents since excess returns would attract new entrants and this increased competition would reduce economic profits and return them to a competitive level. Within the recording industry, it is likely that only certain big name performers earn rents. For most other participants, a decrease in returns would affect their decisions regarding their creative activity. In the absence of rents, any activity that decreases the returns to a supplier or originator of material will result in a reduction in output. If home taping can be shown to have an impact on rates of return or profits, it would thus be likely to affect creative output.

IV. IMPACT OF HOME RECORDING

The focus of the analysis of the impact of audio home taping will necessarily be the recording industry. Within the industry, the major participants likely to be most affected include composers, lyricists, performers, music publishers and the record companies themselves. Although the magnitude of the impact on each of these groups will be different, the direction will be similar if home taping reduces sales since each depends on sales for at least a portion of its income.7

7. The recording artist, record producer, composer, lyricist and music publisher are all normally reimbursed on a fee per copy sold basis. The recording artist's royalties, which are calculated as a percentage of price for each record sold, are often paid only after the initial recording costs have been covered. Statistics Canada estimates that for recording costs of $50,000, an album would have to sell 77,375 copies before artist's royalties would be paid under such an arrangement (cat. no. 87-001, 1980). If the record producer is an independent agent, he will, like the recording artists, be paid a percentage of the price for each record sold. Under the Copyright Act, composers and publishers are reimbursed a set fee for each copy sold. At present in Canada, composers and publishers receive 2¢ per tune for each record sold. An important distinction between composers and publishers on the one hand and performers and record companies on the other is that the former receive performance royalties from broadcasters and other commercial users and thus have an alternate source of income other than record sales royalties.
The following analysis will therefore concentrate on the record industry as a whole rather than on each of its component parts.

For the purposes of this paper, a record company will be regarded as including the recording artists, producers, composers, lyricists and publishers and will be understood to be the agent responsible for the creation, release and promotion of the sound recording.

The revenue received by a record company from the sale of a recording may be written as follows:

\[ R_i = P_i Q_i \]

Equation 1

where \( P_i \) represents the wholesale price per copy sold and \( Q_i \) represents the quantity sold of a particular record.

Within the industry the two major items produced are phonograph records (both long-playing albums and singles) and pre-recorded tapes, which consist in the main of cassettes and 8-track cartridges. In the past, tape cassettes and cartridges have cost more than albums even when the same recorded works were included. Recently, however, this price differential has been narrowing and many companies now have a uniform price for both records and tapes.

The normal practice in the record industry is to fix a list price for the phonograph record or tape, which is also the price appearing in the company's catalogues. As discussed by Blomqvist and Lim (1981), suggested retail prices are extremely uniform in nature. The wholesale price received by the record company will, however, be substantially lower than this list price because of the discounts given to wholesalers and rack jobbers. For example, Statistics Canada has reported that:

During 1977, a typical popular music album with a suggested list price of $7.98 would have been sold by the reporting companies to wholesale distributors or rack jobbers for approximately $4.20 plus or minus 5% depending on the volume of the order. (Statistics Canada, cat. no. 87-001, 1979, p. 2)

8. A rack jobber is a sub-distributor who "racks" or sells records and tapes to retail outlets and maintains control over the purchasing and inventory level of the product for the store.
Thus, there does not appear to be a wide variation in either the discounts given or the wholesale prices received by the record companies for their products. For this reason, the analysis will be simplified by assuming that once the record company has determined its list price for a record or tape, it has also determined the wholesale price (Pi) which it plans to receive per copy.

Turning to the cost side, the cost schedule for a particular recording can be written in the following simple format:

\[ C_i = F_i + M_i Q_i \]  

**Equation 2**

where the \( F_i \) represents the fixed cost\(^9\) involved in creating and marketing the work and the \( M_i \) is the marginal cost per copy made.

The fixed costs component includes the costs of sounds to produce the initial master tape. They include, as well, the promotion and advertising expense incurred in marketing the sound recording and the costs of preparing a pressing or manufacturing run and the cover design.

The marginal costs are those expenses which depend on the number of copies produced. They include the costs of the raw material inputs, such as polyvinylchloride resin or composition for the phonograph records, polyethylene film for wrapping albums and kraft paper for sleeves, required in manufacturing records and tapes.\(^10\)

The profits resulting from an individual recording can thus be written as:

\[ -i = R_i - C_i = (P_i - M_i) Q_i - F_i \]  

**Equation 3**

Home taping may affect a record company's profitability through its impact on sales of pre-recorded material. The existence of low cost high quality home taping equipment

9. For this analysis fixed costs will be defined, as in Blomqvist and Lim (1981), as the one-time costs of adding a new title or record to a company's catalogue. They are not to be considered fixed or immutable over time or identical between different records. The additional costs per extra copy manufactured are the marginal costs.

potentially reduces sales as consumers tend to substitute home-recorded tapes for pre-recorded tapes and records. Not all home recordings are made as direct substitutes for the purchase of pre-recorded tapes or records. Tapers may tape either from their own collection (to preserve the quality of the sound or to have their records on tape to use away from home), from borrowed recordings or off air. Consumers with taping equipment have more flexibility in that they are able to construct music tracks containing selections from many different albums and tapes in ways that were previously impossible. Thus, in many instances, while tapers are willing to tape they are not actually prepared to pay the price to purchase the pre-recorded tape or the record from which it is recorded; taping in such circumstances does not reflect lost sales.

A factor further complicating the analysis of the impact of home taping on the sales of pre-recorded tapes and records is that the existence of affordable, sophisticated home taping equipment is likely to increase the individual's awareness and appreciation of all forms of music. This would have the effect of increasing total consumer expenditure on musical entertainment. Thus, while consumers may be buying home taping equipment and blank tapes, they may also be spending more on pre-recorded material.

This complementarity or sales inducement argument has a number of facets. If one focuses on the increasingly sophisticated home playback stereo systems, including turntables, cassette decks, speakers, radio receiving units, etc., there can be no doubt that the improved technical quality of the sounds has added to consumers' pleasure and enjoyment from listening to music. This, in turn, induces increased expenditure on all facets of this entertainment medium, including pre-recorded material. If the focus is restricted to the actual recording device (which is often included as part of a larger system), then the complementarity or exposure argument is more subtle. The possibility of this type of complement-

11. Many individuals do in fact purchase an album and then transfer the music on to blank tapes to retain a high quality sound. An album may become warped or scratched over time, resulting in a poorer quality sound. Generally the tape in the pre-recorded cassettes sold by the recording companies is thinner and of a poorer quality than many of the higher quality blank tapes on the market. The sounds are also generally poorer than on the phonograph records. This leads many audiophile enthusiasts to buy an album and then transfer it to the higher quality blank tapes.
arity leading to increased income for the recording industry can still exist, however, in a number of ways, including the following:

a) Increased awareness and appreciation of music is very likely to generate spillover effects for performers in the form of increased popularity. This, in turn, will increase both performers' monetary incomes from concerts, endorsements and personal appearances, and their non-monetary incomes or psychic rewards.

b) Individuals may purchase an album jointly and share it for taping purposes, whereas before the advent of home taping equipment none of them would have been willing to pay the full price.

c) The ability to put music together in different formats and to tape one's own albums on high quality blank tapes which retain the sound quality may actually induce some individuals to purchase more albums than before.

d) The cassette recorders now on the market contain both a record and a playback capacity. Thus, individuals who own a cassette recorder will likely purchase both blank tapes for taping and playback purposes and pre-recorded tapes for playback.

From the above discussion, it can be seen that the impact on sales of pre-recorded tapes and records resulting from home taping must be measured as the difference between the substitution and the exposure or complementarity effects. Music taped either off air or from borrowed tapes or records does substitute for the purchase of pre-recorded material. However, the increase exposure and interest resulting from consumer access to home taping equipment makes the overall direction of the impact theoretically ambiguous.12

The extent to which home recording does or does not reduce sales of pre-recorded material is, of course, an empirical question and one which will be returned to later. If, however, the substitution argument is paramount and the quantity of pre-recorded material sold (Qi) falls, then it can be seen from equation 3 that profits will rise or fall depending on the movements in (Pi) and (Mi), since fixed costs (Fi) do not change with changes in output. If neither (Pi) nor (Mi)

12. For a similar discussion of the impact of photocopying on the publishing industry, see S. Liebowitz, 1981.
change, total profits will fall since the product, 

\((B_i - M_i) Q_i\), will fall as \((Q_i)\) falls. Only if the markup of price over marginal cost \((P_i - M_i)\) rises enough to offset the impact on total revenues resulting from the decreased quantity sold will there be an increase in profits.\(^{13}\)

If, on the other hand, the exposure effect dominates, quantity sold will increase and profits will rise unless the markup of price over marginal cost is squeezed to such an extent as to override the beneficial impact resulting from the increased sales.

Some indication of the impact on prices can be deduced. Pre-recorded material provides a convenience factor by making it unnecessary to expend time and energy to tape music. Home taping equipment, however, provides an added measure of flexibility in that individuals are able to assemble the music in arrangements that best suit their own tastes and interests. Assuming no difference in the quality of the sound, the divergence in price between pre-recorded tapes and records and blank tapes should reflect the difference in the relative values consumers place on the convenience and flexibility factors.\(^{14}\) It is clear, however, that the introduction and development of home taping activity has added a further level of competition which needs to be taken into account when record companies are devising their pricing strategies.

In more technical terms, since home-recorded tapes are close substitutes for pre-recorded material,\(^{15}\) the cross-price elasticity of demand will be high for these products.

---

13. Presumably, however, if this was possible the record companies, acting as profit maximizers, would have already been producing at this reduced level.

14. This assumes that consumers own both record players and some form of taping equipment. For those consumers for whom this does not hold, the cost of purchasing the equipment must be taken into account and therefore the price differential between pre-recorded material and blank tapes is unlikely to reflect the difference in the convenience and flexibility factors exactly.

15. The poorer quality of pre-recorded tapes compared to the blank tapes tends to weaken the argument that these two products are close substitutes. This discrepancy in quality increases the substitution away from pre-recorded material.
This would indicate that if the price of pre-recorded material increased, the demand for blank tapes would rise at the expense of the demand for the pre-recorded material. The existence of home taping tends to increase the own-price elasticity of demand for pre-recorded material since a price rise will reduce its demand more than it would have before the existence of home taping. For this reason, the introduction of home taping equipment and the subsequent increased incidence of unrestricted home taping will curtail record companies' ability to increase price (Pi) in the face of increased competition from this source. Therefore, while home taping may cause lost sales, it also prevents recording companies from compensating for any lost sales by raising the prices of pre-recorded tapes and records.

The marginal costs of pressing an additional record or manufacturing an additional tape are unlikely to change dramatically as the quantity produced changes since, for each record pressed, the same amount of raw material input will be required. For this reason, a change in quantity sold is unlikely to change the marginal costs of producing an extra copy to any great extent. If there are any economies of scale in purchasing certain of these material inputs, then a decrease in quantity sold should increase the marginal costs since, at lower scales of output, the company might be unable to realize such economies.

Home taping will therefore exert some downward pressure on prices (Pi), while its impact on quantity sold (Qi) is more ambiguous. Fixed costs (Fi) and, in all probability, marginal costs (Mi) will remain constant. It has been seen from equation 3, - i = (Pi - Mi) - Fi, that the net result of these effects on - i - the profits from the sale of records - depends crucially on the net impact on quantity sold. While the magnitude of the impact of home taping on record sales and prices is likely to differ for different companies and across the range of classes of records, the direction of the effect should be similar in all cases.

Summary

In this section, the effect of home taping on record industry profits, including the returns to composers and performers, has been outlined. It has been argued that home taping could potentially affect both the price and quantity of pre-recorded tapes and records sold and, as a result, the total revenues from the sale of these goods. Any negative impact on sales would probably result in decreased profits due to the need for large sales volumes to offset the high fixed costs of creating a new recording. This impact is important from a copyright perspective since profits provide the economic incentive to create.
V. SUMMARY OF HOME TAPING PRACTICES

Surveys attempting to assess the extent of private home copying of music have been conducted in almost every major western country. Space limitations do not allow us to summarize in detail the findings of these surveys or to make international comparisons. In this paper, therefore only the highlights of a recent Canadian survey [6] will be reported upon in order to demonstrate the impressive growth of this activity during the 1970's.

In 1971 about 25 per cent of Canadian homes owned some type of recording equipment. By 1981 this had risen to more than 60 per cent. Usage of the machines to tape copyrighted works is quite high as 76 per cent of those who own such equipment claim to tape material which normally is protected by copyright during the year prior to the conducting of the survey. The volume of taping undertaken by the home tapers has also been increasing.

The highest percentage of people tape from either their own or borrowed records or tapes. The reasons most often cited for why people tape are that 1) it allows them to create their own selection of works; 2) it is very easy to play these tapes and they are easily portable for use in car or cottage tape decks and; 3) to save money. These reasons for taping helps to explain why it is that heavy tapes also tend to be the heavy purchases of pre-recorded records and tapes.

VI. Canadian recording industry performance

This section charts the performance of the Canadian recording industry in terms of record releases, prices and sales as these were the elements identified in the theoretical section as most likely to be affected by home taping. In Canada, the vast majority of pre-recorded tapes and records sold are manufactured domestically. Statistics Canada estimated that for 1978 the imports of phonograph records equalled only 8.6 per cent of the total sales revenue earned by record manufacturing companies (Cat. no. 87-001, 1980, p. 26).

The percentage of the total market accounted for by imports has therefore remained relatively constant, as it was estimated at eight per cent for 1971 (Klopchic, 1976, p. 25).

16. The survey results reported on in this section of the paper are from a survey carried out by Market Facts of Canada Limited on behalf of Consumer and Corporate Affairs Canada. A comprehensive summary of the survey results is contained in Keon 1982 (b).
An examination of the figures for number of domestic releases and net shipments of records and tapes manufactured will thus provide an accurate picture of total record industry activity in Canada.

Table I demonstrates the growth in both the number and value of net shipments of records and tapes in Canada. From the table it can be seen that throughout the 1970's and especially after 1975 until 1979 the rate of increase in sales was impressive. The figures for 1979-1981 demonstrate the current record industry slump. Sales volume stagnated in 1979, fell off 10 per cent in 1980 and the 1981 level is seen to be approximately 9 per cent below the 1978 level.

This slump in sales for the recording industry is paralleled by similar declines in most other developed countries. Not surprisingly the call for compensation for the recording industry as a result of home taping has become stronger during this recent decline.

The growth in sales throughout the 1970's did not simply reflect population growth. Table II indicates that for 1979 more records and tapes were sold per household than ever before. The figures for 1980 and 1981 reflect the downturn in sales indicated in the gross shipments figures.

17. Statistics Canada defines releasing a record as the arranging for the pressing of discs or the duplication of tapes in either company-owned facilities or by a custom pressing contract with a pressing plant. Thus, both domestically created and imported master tape recordings are considered to be released in Canada when phonorecords or tapes are manufactured from the master tape (cat. no. 87-001).

18. Statistics Canada defines net shipments as gross shipments, less returns and exchanges. Net shipments do not include promotional records shipped to radio stations or other reviewer samples (cat. no. 47-004).

19. For an interesting if slanted discussions of the unlikelihood that home recordings is responsible for the ills of the record industry slump see Robert Wallace 1980 and Michael Schrage 1982.
| Table I |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Net Shipments of Records and Tapes Manufactured in Canada, 1970-1981** |
| **Millions of Units** | | | | | | | | | | | |
| Phonograph Records | | | | | | | | | | | |
| 7" | 15.5 | 14.9 | 16.8 | 14.2 | 14.8 | 14.6 | 18.1 | 19.8 | 17.3 | 20.7 | 16.7 | 14.1 |
| 12" | 25.0 | 26.4 | 29.4 | 30.1 | 32.0 | 31.8 | 38.9 | 46.0 | 53.9 | 53.9 | 52.4 | 54.4 |
| Total Records | 40.5 | 41.3 | 46.2 | 44.3 | 46.8 | 46.4 | 57.0 | 65.8 | 71.1 | 74.6 | 69.1 | 68.5 |
| Pre-Recorded Tapes | | | | | | | | | | | |
| 8-track cassettes | 2.9 | 3.2 | 5.9 | 7.1 | 9.7 | 11.2 | 15.6 | 15.0 | 15.7 | 10.9 | 5.2 | 2.5 |
| Total Tapes | 3.6 | 4.1 | 6.9 | 8.2 | 10.8 | 12.7 | 18.4 | 19.5 | 32.0 | 19.9 | 15.4 | 17.3 |
| Grand Total | 44.1 | 45.4 | 53.1 | 52.5 | 57.6 | 59.1 | 75.4 | 85.3 | 94.1 | 94.5 | 84.5 | 85.8 |
| % Change (1) | - | 2.9 | 17.0 | -1.1 | 9.7 | 2.6 | 17.6 | 13.1 | 10.3 | 0.4 | -10.6 | 1.5 |
| **Millions of Dollars** | | | | | | | | | | | |
| Phonograph Records all sizes | 49.9 | 57.5 | 63.7 | 73.6 | 92.7 | 100.3 | 112.9 | 139.3 | 188.5 | 215.1 | 199.8 | 107.3 |
| Pre-recorded Tapes all formats | 15.8 | 17.0 | 24.9 | 31.5 | 39.0 | 40.3 | 47.2 | 52.1 | 70.0 | 63.1 | 53.3 | 61.5 |
| Total | 65.7 | 74.5 | 88.6 | 105.1 | 131.7 | 140.6 | 160.1 | 191.4 | 258.5 | 278.3 | 253.1 | 268.8 |
| % Change (1) | - | 13.4 | 18.9 | 18.6 | 25.3 | 6.8 | 13.9 | 19.6 | 35.1 | 7.7 | -9.1 | 6.2 |

(1) Percentage change from previous year.

Table II

Disc and Tape Sales per Household, 1970-1981

<table>
<thead>
<tr>
<th>Year</th>
<th>Discs</th>
<th>Tapes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>7.2</td>
<td>0.6</td>
<td>7.8</td>
</tr>
<tr>
<td>1971</td>
<td>7.1</td>
<td>0.7</td>
<td>7.8</td>
</tr>
<tr>
<td>1972</td>
<td>7.6</td>
<td>1.1</td>
<td>8.7</td>
</tr>
<tr>
<td>1973</td>
<td>7.1</td>
<td>1.3</td>
<td>8.4</td>
</tr>
<tr>
<td>1974</td>
<td>7.2</td>
<td>1.7</td>
<td>8.9</td>
</tr>
<tr>
<td>1975</td>
<td>6.9</td>
<td>1.9</td>
<td>8.8</td>
</tr>
<tr>
<td>1976</td>
<td>8.2</td>
<td>2.7</td>
<td>19.9</td>
</tr>
<tr>
<td>1977</td>
<td>9.4</td>
<td>2.8</td>
<td>12.2</td>
</tr>
<tr>
<td>1978</td>
<td>9.7</td>
<td>2.7</td>
<td>12.4</td>
</tr>
<tr>
<td>1979</td>
<td>9.9</td>
<td>2.6</td>
<td>12.5</td>
</tr>
<tr>
<td>1980</td>
<td>8.9</td>
<td>2.0</td>
<td>10.9</td>
</tr>
<tr>
<td>1981</td>
<td>8.5</td>
<td>2.2</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Source: Calculated by dividing net shipments data from Statistics Canada, Cat. No. 47-004, by the number of households in Canada as reported in Statistics Canada, Cat. No. 64-202.

In an earlier section of this paper it was stated that the output of the recording industry in terms of creative production should be measured by the number of master tapes produced and released since this is the most accurate measure of creative activity. Table III presents figures for the number of releases by the Canadian recording industry for certain years.
Table III

Number of Records Released

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Singles</td>
<td>1618</td>
<td>1331</td>
<td>2310</td>
<td>2031</td>
<td>1,401</td>
</tr>
<tr>
<td>Albums</td>
<td>1780</td>
<td>1717</td>
<td>3148</td>
<td>2700</td>
<td>1,885</td>
</tr>
</tbody>
</table>

Source: The figures for 1970 and 1971 are taken from a special edition of Statistics Canada survey, Cat. No. 47-004 (Production and Sales of Phonograph Records and Pre-Recorded Tapes in Canada) which was subsequently discontinued. The latter figures are from the recently originated Statistics Canada cultural industries survey 87-001.

The quantity of new recordings released exhibited an impressive growth pattern from 1971 until 1977 growing at a compound rate of approximately 8 per cent per year for new albums. However, figures for 1978 and 1979 show a dramatic decrease falling off 14 per cent in 1978 and a further 30 per cent in 1979. Thus in 1979 there were fully 40 per cent fewer albums released than in 1977. Unfortunately at the time this manuscript was prepared data for more recent years were not available and so it was impossible to check on whether this dramatic decline in new album releases has continued.

In the theoretical section, it was also argued that home taping could potentially restrain increases in prices of pre-recorded material. Table IV shows comparison of the retail price of records with all other elements of the consumer price index including food.
### Table IV

**Consumer Price Index for Records and All Non-Food Items, 1971-1981**

<table>
<thead>
<tr>
<th>Year</th>
<th>Consumer price index for records, 1971 = 100</th>
<th>Consumer price index for Canada -- all items excluding food, 1971 = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual average</td>
<td>Annual average</td>
</tr>
<tr>
<td>1971</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>1972</td>
<td>113.8</td>
<td>103.7</td>
</tr>
<tr>
<td>1973</td>
<td>119.3</td>
<td>109.9</td>
</tr>
<tr>
<td>1974</td>
<td>137.5</td>
<td>118.6</td>
</tr>
<tr>
<td>1975</td>
<td>146.3</td>
<td>130.5</td>
</tr>
<tr>
<td>1976</td>
<td>147.4</td>
<td>142.8</td>
</tr>
<tr>
<td>1977</td>
<td>154.2</td>
<td>154.0</td>
</tr>
<tr>
<td>1978</td>
<td>Sept. 158.7</td>
<td>163.9</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td>176.9</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td>194.6</td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td>219.5</td>
</tr>
</tbody>
</table>


*The price index for records is for popular large volume selling albums only.*

At the retail level, the price index for records initially increased faster than the price index for all items excluding food. The spread in the indices narrowed after 1975 and, as of 1978, they were approximately equivalent. Unfortunately Statistics Canada's confidentiality requirements do not allow the publishing of a record price index beyond this point. The unpublished data do, however, indicate that the retail price of records has risen at a slower rate than the general rate of inflation as measured by the CPI.

In an attempt to correlate more closely the data on home taping activity with the performance of the recording industry in Canada it was decided to formulate and test the following regression equations. Data regarding the growth in both home taping equipment sales and the sales of pre-recorded tapes and records have been reported. Lack of accurate data on sales of home taping equipment for a sufficiently long
period of time prevented the determination of a precise statistical relationship between home taping and record sales. Nonetheless, an attempt was made to use the available data to the extent possible. The central reason for attempting the regressions was to establish the impact of home taping on record sales in Canada. A short description and rationale for the variables to be included in the regression analysis will be given. The data and regressions cover the period from 1971 to 1981 inclusive.

The first determinant of demand for records is their price since demand varies inversely with price. Thus, the sign of the price variable should be negative. The next determinant of demand for record sales is level of income. For records and indeed for almost all commodities, the greater the money income the greater the demand. The income variable should, therefore, have a positive sign. The remaining variables to be included in the equation are the prices of related commodities considered in this paper. Record playing equipment is a complementary item. Thus, if the price of record playing equipment falls and more individuals purchase this equipment, demand for records should increase. Home taping equipment, on the other hand, may represent a substitute good. As explained in the text, individuals will, in certain instances, tape music either off air or from borrowed records or tapes in place of purchasing the pre-recorded material. Under these circumstances, a decrease in the price of home taping equipment with a concomitant increase in demand for it should result in a decrease in demand for pre-recorded tapes and records.

The problem with this construction is the lack of data on the price of either record players or home taping equipment for Canada. Changes in existing stocks of record players and home recorders (for which data are available) can, however, be used as proxies for their prices since it can be argued that there is a direct inverse relationship between the two.

There are some potential biases in this approach in that it can arguably be stated that the demand for record players and home taping equipment might be functions of the price of pre-recorded tapes and records. If the price of these latter items rises, individuals could have an added inducement to purchase home recorders and tape their own music. Survey evidence presented in this paper has indicated, however, that for the majority of home tapers the critical factor in their decision to purchase home recorders is not the ability to save money by taping music rather than buying the pre-recorded material. Rather, the determining factor was the increased flexibility in terms of constructing one's own tapes. Thus, the price of records does not appear to be the
central determinant in the overall decision regarding the purchase of home taping machines. It is more likely to affect how often and for what purposes the machine is used.

Another variable which would have been useful to the regression analysis would have been the number of new releases. Record sales in year t should depend on the level of new releases in each of the previous years. The propensity to buy a recent release would be expected to be much higher than the propensity to buy an older record. Hence current sales would be expected to decrease if there is a decrease in the number of new releases. This variable could not be included in the regression, however, since as was indicated in Table III above, data for this variable were not available from 1971 through 1976.

It might also have been useful to have included a variable measuring quality of new releases since variations in sales would also be influenced by the quality of new releases. However, it is not clear conceptually how this factor could be best measured. Even if a clear conceptual measure could be formulated the necessary data to test its impact would not be available.

In estimating industry sales it would also seem relevant to incorporate industry advertising expenditures. For the recording industry, however, such data are not useful. This arises since the best advertising record companies receive, airplay of their records, is free. Further advertising which is undertaken is often done at the retail level by individual stores which receive discounts on record purchases from the record companies. These discounts do not show up as record industry advertising expenditures. For these reasons actual advertising expenditures vastly underestimate the value of advertising for the industry and thus their inclusion in the regression would be misleading.

On the basis of these considerations the regressions, which were run using the ordinary least squares technique, took the following form. A description of the precise method of constructing the variables included is given at the end of this section. The regressions were run from the period 1971 to 1981.

\[ RS = a_0 + a_1 P + a_2 Y + a_3 RPE + a_4 HTE \]

where

- \( RS \) = record sales per capita in units
- \( P \) = retail sales price index for records
- \( Y \) = net national income per capita
RPE = number of households in Canada with record playing equipment
HTE = number of households in Canada with home taping equipment

The results of the regressions, which were run in normal and logarithmic format, were as follows:

(1) \[ RS = 1.34 - 0.760P + 0.899Y + 0.736PPE - 1.41HTE \]
\[ \quad (1.709) \quad (1.09) \quad (0.04) \quad (1.27) \]
\[ R^2 = .941 \quad F = 11.65 \quad D.W. = 1.78 \]

(2) \[ RS = -14.50 - 0.571P + 1.1924 + 0.579 - 1.315 \]
\[ \quad (1.15) \quad (1.26) \quad (0.03) \quad (1.09) \]
\[ R^2 = .937 \quad F = 10.83 \quad D.W. = 1.56 \]

The numbers in brackets are the "t" values.

All the variables have the expected signs. However, none of the variables are significant at the 95 per cent level.

The major reasons for the lack of strong results are: a) the small number of observations for each variable (eleven); b) the lack of precise data for sales of home recording equipment; c) the absence of potentially important variables due to lack of data; d) the very high degree of collinearity between many of the independent variables; and e) the lack of significance of the resulting coefficients, perhaps due to the points just raised as well as the possibility that the variables are indeed insignificant.

It is interesting to note, however, that the sign of the home taping variable, though statistically not significant, is negative in both equations. This result would lend some weak credence to the substitution argument (i.e., that people with home recording equipment will tend to substitute home recorded material for pre-recorded tapes and records). In the absence of more accurate and longer term data, it is impossible to go beyond these somewhat tentative conclusions regarding the regression results.

Dependent Variables

Record sales: per capita pre-recorded tape and record sales in Canada in units (source: Statistics Canada, cat. no. 47-004).

Independent Variables

Consumer price index for records: This variable was deflated by the general consumer price index to express it in real terms. (source: Statistics Canada, cat. no. 62-010).
Real net annual income (Y): These figures are per capita annual income (source: Statistics Canada, cat. no. 13-201). The annual figures were deflated by the consumer price index. An age group variable was tested in the regressions but showed no significance whatsoever and thus was dropped.

Record playing equipment (RPE): number of households owning record playing equipment, 1971-81 (source: Statistics Canada, cat. no. 64-202).

Home taping equipment (HTE): number of households owning tape recording equipment (source: Statistics Canada, cat. no. 64-202). These data are collected by Statistics Canada only on a bi-annual basis. Estimates for the intervening years were produced by extrapolation from the bi-annual figures.

VII. IMPACT OF IN-HOME RECORDING: CONCLUSIONS

In the detailed analytical section, alternative consequences of home taping were outlined. Depending on whether the substitutability factor (taping music rather than buying) or the complementarity or exposure factor (interest in music being piqued by the increasingly sophisticated playback and taping equipment, causing purchase of more pre-recorded tapes and records) was preeminent, net sales of pre-recorded music would either rise or fall. It was left to the empirical analysis to come to a conclusion as to what the overall effect has been.

Commissioned survey results on home taping indicated that the ownership of home audio recording machines has dramatically increased over the course of the 1970's and early 1980's. Usage of the machines is quite high as 76 per cent of those who own such equipment claim to tape material which normally is protected by copyright.

The reasons most often cited for why people tape are that it allows them to create their own selections of works and it is very easy to play these tapes and they are easily portable for use in car or cottage tape decks. Taping to save money is claimed to be of lesser importance.

These reasons for taping probably hold the explanation to how it is that the tremendous growth in the home recording of copyrighted musical works throughout the 1970's was able to co-exist with consistent, though less dramatic, annual increases in the volume of sales of pre-recorded records and tapes.
The first major stage of the empirical analysis consisted of an examination of the performance of the Canadian recording industry since 1970. Performance was charted in terms of record sales, prices and number of releases (the elements identified in the analytical section as most likely to be affected by home taping). The results from the early 1970's appear to fit best with the complementarity/exposure/sales inducement arguments.

The data for the late 1970's, however, indicate that the performance of the recording industry has been stagnant while sales and usage of blank tapes have continued to flourish.

The regression analysis failed to produce statistically significant results. The negative coefficient on the home taping variable does lend some credence to the argument that home taping is reducing record sales. The regression analysis has been limited by the lack of accurate data especially on the sales of blank tapes and home recording equipment. As a result, a more precise determination of the impact of home taping of music on sales of pre-recorded music has not been possible.

Despite these limitations in the analysis we conclude that both the weight of the theoretical arguments and the empirical analysis would tend to support the position that home recording is contributing to a decline in the demand for pre-recorded music. It is clear that ownership and usage of home audio recorders has been growing quickly and is likely to continue to do so for the foreseeable future albeit at a much slower rate. Given this and given the negative impact which such home taping can have on the recording industry, analysis of potential compensatory schemes to reimburse the industry for lost revenue is called for.

VIII. ANALYSIS OF POTENTIAL COMPENSATORY MECHANISMS

In assessing the in-home taping issue the initial question to be determined is whether or not any provision for reimbursement for such activity is called for. If it is decided that reimbursement should take place, the next question concerns the form of the reimbursement mechanism.

Determining the basis of the payment scheme not only settles the question of the size of the royalty payments to be collected but can also influence considerably the decision regarding the overall desirability of introducing a payment liability. It would not appear to further the well-being of copyright owners to grant or attempt to enforce a right if a suitably efficient collection system cannot be devised and implemented. Without a workable payment scheme, the entire exercise becomes academic.
There are very serious problems in attempting to apply normal methods of enforcement or collection to compensation for in-home taping (i.e., those methods which entail payment or collection from the actual users of the material). Most previous commentary regarding the collection of royalty payments for in-home taping has argued that requiring payment by individuals on a per use basis for recordings made in the home would be both intractable and undesirable. It would be intractable due to the extreme difficulty and unlikelihood of anyone ever being able to monitor such taping effectively with current technology. In addition, and perhaps more importantly, an individual payment format would be undesirable given the degree to which it would entail the invasion of the individual's privacy at home, especially if technological monitoring is not feasible. Such arguments appear sound and it is unlikely that strong disagreement would be voiced against them. This intractability in formulating a user's payment mechanism had led most compensation advocates to favour a system which would impose liability for payment on the manufacturers or importers of either or both of the home taping equipment and the blank tapes.

Until recently, the Federal Republic of Germany was the only country to have enacted such a system for the specific purpose of reimbursing copyright owners for losses in sales due to in-home audio and video taping. Recently, however, Austria passed legislation which took effect January 1, 1981 and provides for similar compensation and Sweden has passed legislation to tax blank tapes beginning in September, 1982, the basis for an analysis of the strengths and weaknesses of schemes to impose payment liability on manufacturers or importers of home taping equipment. Such analysis is called for.

since these systems, or some close variation of them, are advocated by many copyright associations around the world.21

A Levy on Tape Recorders vs. A Levy on Blank Tapes

As early as 1954, certain German tape recorder manufacturers agreed amongst themselves to pay a copyright remuneration to German copyright holders whose music was being taped at home by means of the manufacturers' equipment.22 This was somewhat surprising because not until 1965, under section 53(5) of the German Copyright Law, was there any legal obligation for them to make such payments.

The essence of the German Act is that it provides for a levy to be imposed on all tape recorders suitable for making recordings of protected works for private use. A proposal to impose a levy on the blank recording tapes was considered but rejected. One of the reasons given for rejecting a blank tape tax was that it was impossible to tell whether the tapes would be used just for dictating, or for recording copyrighted works. With regard to the tape recorder itself, it was felt likely that recording equipment suitable for private taping would at some point during its lifetime be used in that capacity.

This rationale for a levy on tape recorders but not on blank tapes is not entirely convincing. As discussed, tape recorders can both record and play back copyrighted material. Thus, the machines will simultaneously stimulate pre-recorded sales through their playback capacities and retard sales due to their ability to record music on blank tapes. The blank tape, however, has a more one-dimensional effect. It cannot directly stimulate sales of pre-recorded tapes or records. Rather it is seen as a direct competitor or substitute for

21. It has also been suggested that another method for recompensing copyright owners for the private home recording of their works would be to increase the level of performing royalties paid by broadcasters since a large amount of home taping is done off air. This, however, is not the most equitable solution since broadcasters are already paying composers and publishers for the right to use their music and have little control over whether individuals decide to home record or not. In Canada, there is no performing right in sound recordings. For a discussion of the many aspects pertaining to the introduction of such a right see Keon, 1981.

22. For a more complete description of the German system, interested readers should see Reinbothe, 1979.
these goods. For this reason, it would appear to be a more natural candidate for a levy or tax to reimburse copyright owners.

A tax on blank tapes also more accurately reflects use patterns. Individuals who engage in extensive home taping will presumably also purchase large numbers of blank tapes. Thus, a greater burden would be put on those who record the most. This appears to be most in line with traditional copyright payments, which are based on the popularity or sales of the work. It would not result in an exact matching of use and payments since many tapes are used more than once, but it would still result in compensation more in line with use than would a tax on hardware alone.

The Whitford Committee was of the view that a levy on blank tapes would involve a much larger operation and would therefore be most costly administratively (paragraph 308). While no direct evidence is available on this question, a levy involving a fixed amount per tape imposed on either the manufacturer or importer of blank tape should not prove too complex or costly. The benefits, in terms of more accurately reflecting actual use, would not appear to be outweighed by these cost considerations.

The Austrian and Swedish legislatures appear to have reached similar conclusions regarding the desirability of a levy on hardware vis-à-vis software: their new legislation imposes a statutory levy on blank tapes.

Unavoidably there will be many inequities arising from a blank tape levy. Blank tapes are used for many purposes in addition to taping copyrighted material. They are frequently used for office dictation and for amateur home recordings, which have nothing to do with copyright. It is not presently known what proportion of blank tape sales in Canada is accounted for by commercial or business establishments, who would rarely if ever use them for taping copyrighted music or programs, and what proportion is accounted for by home purchasers. The survey results reported on in this paper indicated that of all households owning a tape recorder approximately 76 per cent had taped copyrighted material on a blank tape in the past 12 months. It is obvious that a very large proportion of blank tapes sold will never be used to record copyrighted works.

It would be practically impossible to identify precisely the intended use of a tape when it is purchased. There does not appear to be a foolproof method of imposing a levy that will not result in inequitable treatment for some blank tape purchasers. If the decision is taken to impose a blank
tape levy then inequities such as imposing a tax on tapes which will not be used for home taping will have to be accepted. There are many inefficiencies in this approach but it is preferable to attempting to capture funds each time an item is taped at home. If a blank tape levy were introduced, manufacturers of these goods should be allowed to present information regarding the expected use of their particular brands of tape. If differences in use can be identified precisely, the levy should be adjusted accordingly.

Clearly, the manufacturers and importers who might be required to initially pay the tax would have an incentive to shift the tax on to purchasers. The extent to which they are able to do so would transfer the burden, or the "economic incidence", of the tax away from its initial impact point. With regard to a blank tape tax, such a shifting should not cause undue concern since the case for placing the burden of collection at the manufacture and import level was made on the basis of convenience and administrative ease. It is the individuals making use of the copyrighted material who have traditionally paid copyright owners and, to the extent that the incidence of the blank tape tax is shifted, this would be in line with these traditional concepts.

Whether manufacturers and importers are able to shift the tax depends on the elasticities of supply and demand. The more inelastic the demand curve with respect to price, the more the manufacturers and importers will be able to shift the burden of the tax to consumers. By contrast, the more inelastic the supply curve, the greater the burden on the manufacturers and importers.

The major factors determining the elasticity of demand are the availability of substitute products and the importance of the product in the consumer's overall budget. As argued in Section IV, blank tapes and pre-recorded copyrighted material are substitute products. This indicates that the demand curve for blank tapes has a degree of elasticity. Given that there is competition among manufacturers of blank tapes, it is reasonable to assume an infinitely elastic supply curve. The consequence of these factors is that if manufacturers and importers attempted to increase price by the full amount of the tax, there would be some consumer resistance, and in the aggregate, there would be fewer blank tapes sold.

The exact amount by which sales of blank tapes would fall depends on the interaction of the demand and supply curves. A precise estimate of the shifting of the tax or the decline in sales of blank tape as a result of a levy would require exact estimates of the demand and supply elasticities. No such estimates exist or will be attempted here.
This discussion has highlighted the fact that the imposition of such a levy will result both in consumers paying more and in the amount of blank tapes being purchased falling.23

The competitive position of pre-recorded tapes and discs vis-à-vis blank tapes will increase. Thus, copyright owners should benefit in two ways from the establishment of a blank tape tax. They will reap the revenues resulting from the levy itself directly. In addition, to the extent that the levy increases the saleability of pre-recorded tapes and discs (from which copyright royalties are paid), the owners will benefit from increased royalties from this source as well.

Method of Calculation and Level of Royalty Payments

Two bases can be used to determine a format for calculating the tax levy: a) a percentage of price system, either at the wholesale or retail level, or b) a flat rate levy per tape.

A flat rate levy per tape is preferable in that the same music can be taped on blank tapes regardless of their price and it is the taping of music in lieu of purchasing for which a reimbursement is being paid. The IFPI home taping policy statement arrives at the same conclusion for the following reasons.

Having studied both these systems, IFPI has no doubt that (b) is the more equitable. Neither the manufacturer's price nor the wholesale nor the retail price of blank tape should be used as a basis for calculating the royalty, because the price of the tape is irrelevant to the use to which it is put or the gain to the consumer. The tape is only the material support for the music or other work recorded thereon. It is the work, not the tape, which determines whether and to what extent there is damage caused to producers of phonograms, performers and authors. The damage inflicted by the private copier does not depend on the price of the blank tape (there are substantial differences in price between low

23. For a mathematical exposition of the impact of a sales tax in a competitive market, see J.M. Henderson and R.E. Quandt, 1971, pp. 124-126.
quality and high quality blank tape) but depends on the playing time of the cassette. Therefore (b) provides a fairer basis. The royalty would be calculated as an "equitable percentage" of the royalties and return indicated in (b) above and would be collected from manufacturers and importers of blank tape. (IFPI, 1979, p.6).

Since blank tape can have different time recording capacities, the optimal calculation method should be based on a standard rate per unit of time (i.e., so many cents or fractions of cents per minute).

Method of Collection and Distribution

Administratively, the most efficient means for collecting the royalties would appear to be along the lines of the German system. Royalties are collected from the manufacturers and importers since there are fewer of them than there are retail establishments which sell these items. Liability must be placed on importers since it would be impossible to enforce a claim against foreign manufacturers.

Such a system would place an obligation on manufacturers and importers to supply information to a collecting society regarding shipments and sales of blank tape by them. Since the copyright interests will benefit from a compensation scheme, it would seem only fair that these interests be responsible for collecting the levy at rates subject to the approval of a copyright tribunal. This approach is currently used to collect performing rights royalties in Canada and there does not appear to be any good reason to depart from this practice.

After the level and method of collecting a royalty have been decided upon, the next step is to develop a practical method for distributing the royalties. In Germany a central agency referred to as ZPU24 was founded jointly by the three associations representing authors, composers of music and performers and producers of sound recordings who are entitled to compensation under the German scheme. The distribution of the money collected by ZPU is carried out in two steps. First it is distributed to the three agencies and then they distribute it to their members in line with their particular relevant statutes. The major difficulty arises from the fact that

24. Zentralstelle fur private Überspielungsrechte, the central agency for private recording rights.
it is not possible to tell which works are being recorded or how many times they have been recorded.

The frequency of radio and television air play, the magnitude of sales of a work or some combination of the two would seem to constitute the most appropriate method of calculating royalties for musical works for distribution purposes. In any event, the decision regarding the apportionment of royalties under a copyright type scheme should be left to the societies themselves according to the agreed-upon distribution rules of the members of the society. Again, this is presently the practice with the performing rights societies in Canada and there does not seem to be a compelling case for further government intervention at this level.

Decisions regarding the distribution of royalties among the major groups would, however, presumably require a body such as a Copyright Royalty Tribunal. Such decisions should be made on the basis of the relative or proportional returns of these groups from the sale of pre-recorded material. If such a scheme were introduced, information would have to be presented to allow a Tribunal to make its decision in an enlightened manner.

An important factor to remember when discussing the distribution of royalties to authors, performing artists and producers under a copyright formula is that, in Canada, performances do not qualify for copyright protection. It would be incongruous, therefore, to distribute royalties generated by a levy scheme operating under the auspices of the Copyright Act to a group which does not qualify for copyright protection. A scheme not within the purview of the Copyright Act could, without such an incongruity, make such payments to performing artists.

In addition, a non-copyright oriented scheme would not require the payment of royalties to non-Canadians or could allow for payments only on a reciprocal basis. Canada's international copyright obligations prevent payments from being made in such a discriminatory fashion. The present German system operates under their copyright law and gives rise to national treatment. Whether a Canadian scheme should operate outside the ambit of copyright legislation can only be decided after taking into account economic and political considerations, such as the outflow of royalty payments and the potential size of foreign payments to Canadian artists.

IX. CONCLUSIONS - RECOMMENDATIONS

Based on the analysis of this paper it is suggested that a levy on the sale of blank tapes to reimburse Canadian
composers, performers and record companies for lost revenue occasioned by home recording can be justified. As indicated in the last section careful consideration needs to be given to mechanisms of collecting and distributing the levy in order that it operate in the most efficient manner.

It is further suggested that if a levy scheme is introduced in Canada it should be done so initially outside the ambit of the Copyright Act. Removing this system from copyright allows the Canadian authorities to structure it to the benefit of the indigenous elements of the Canadian recording industry. If and when other countries with significant recording industries institute similar systems, consideration could be given to extending the benefits of the Canadian levy to foreigners on a reciprocal basis.

Concurrent with the imposition of a levy an exemption legitimizing the private non-commercial home recording of music should be introduced into the Copyright Act. Such an exemption is required since it is clearly undesirable to have the Copyright Act being constantly breached by millions of people recording at home.
BIBLIOGRAPHY


COPYRIGHT AND PHOTOCOPYING:

ALTERNATIVE INSTITUTIONAL ARRANGEMENTS

S. J. Liebowitz
Department of Economics
University of Western Ontario
London, Canada
Studies undertaken during the recent phenomenal growth of photocopying have indicated that the largest subset of photocopied items is scholarly journals. Publishers of journals increasingly have been engaging in price discrimination. The ramifications of these evolving forms of behavior are not well understood by authors, publishers or the legal profession. This is most unfortunate since the impact of reprography on copyright is a topic of governmental concern in most Western countries. This issue has pitted publishers and authors against librarians in the debate over changes in copyright law recently accomplished in the U.S. and contemplated in Canada.

In this paper price discrimination will be shown to influence the working of copyright legislation in ways heretofore unsuspected. Price discrimination circumvents the concept of fair use and imposes a public lending right even where none is legislated. In addition it can, in principle, be a perfect substitute for strict copyright protection in the making of photocopies.

At issue is the question of the proper degree of copyright protection which should be extended to the making of photocopies. We shall examine the rationale for copyright protection in principle and then shall compare the various necessarily imperfect institutional arrangements which attempt to provide intellectual property protection. Empirical investigation will reveal the apparent superiority of market arrangements (price discrimination) over those imposed outside the market (the Copyright Clearance Center).

1For a summary of these studies see Liebowitz [1980a].
1. **Copyright Law**

Copyright gives authors certain property rights over their intellectual creations. Most importantly, it gives copyright holders the sole right to reproduce or publish the work. This right has been extended in various ways so that it now covers public performance, translation, adaptation and broadcasts of the copyrighted work in most western countries. Various international agreements ensure minimum levels of protection in the signatory countries. Copyright protects the expression of intellectual ideas, not the ideas themselves. Independent creation of the same work is not prohibited. The quality and purpose of the intellectual material is irrelevant in the granting of copyright. The usual term of copyright is the life of the author plus fifty years.

There are various exceptions to copyright protection such as performance for charitable causes or use of short passages in schools. The most important exception, which has been of great importance to academic and other researchers making photocopies (perhaps unbeknownst to them), is known as fair use (sometimes called fair dealing). Fair use is a defense to a claim of infringement currently provided in Anglo-American copyright law. It is stated in section 17(2)(a) of the Canadian Law that "The following acts do not constitute an infringement of copyright: (a) any fair dealing with any work for the purpose of private study, research, criticism, review or newspaper summary." Section 107 of the new U.S. act (which took effect in 1978) has very similar phrasing and excepts fair use for purposes of criticism, comment, news reporting, teaching, scholarship and research. The courts determine whether a particular action constitutes fair use and no hard and fast demarcation exists.
One right which is not granted under U.S. or Canadian law is a public lending right. Such a right would entitle authors to compensation whenever their copyrighted works are borrowed from libraries. West Germany, Sweden, Denmark, Australia and Norway are among those countries already granting public lending rights. The United Kingdom is seriously considering such a law.  

Copyright is only one of several possible methods whereby authors or publishers can appropriate revenues from those who use intellectual properties. Arnold Plant 3 claimed that being first in the market allowed authors to capture a good deal of the potential revenue and there is no doubt that there is some truth in this belief although modern technology has reduced the time advantage of first publication. Another way by which revenues can be appropriated is that of discriminatory pricing. If those who plan to make competing editions of an intellectual property can be charged a very high price in their purchase of an original, the original producer can capture the value from those consumers of the competing edition.

2. Technology and the Need for Copyright

The rather unusual aspect of intellectual property, which reduces the author's control of the property and necessitates 'special' property rights is its public good characteristic. My consumption of a poem, or an idea,

---


or a song does not reduce anyone else's possible consumption of these items. The public good aspect of intellectual property is not by itself sufficient to require any special property right such as copyright. It is also necessary that technology be such that physical representations of the intellectual property can be made. When the intellectual property can be disembodied from its physical representation the dilemma which we have come to associate with intellectual property arises. This dilemma exists because the exclusion of people from consuming the intellectual property becomes impossible unless one can have complete control over the creation of the physical representations of the intellectual property. The physical representations of intellectual property (e.g., the paper on which it is written) have private good characteristics such that my reading of a book precludes others from reading the same book at the same time. If one cannot control the creation of physical representations of an intellectual property then others will produce physical representations of the intellectual property (assuming there is no law forbidding such behavior) and one would be unable to exclude people from using the intellectual property.

A necessary condition for the market to effectively produce goods is that it must be possible to exclude people from using the good. No one would pay for an automobile if he could use any car he wanted without paying for it. Property rights over private goods provide exclusion in most cases since the law will prosecute those who use the good without transacting for it.

The role of technology in disembodying the intellectual property from its creator must not be underestimated. We are all so familiar with the printing press that it becomes difficult to imagine authors not being disembodied from their work. However, prior to printing presses authors had
to tell their tales in person. Only someone with an exceptional memory would have been able to reproduce the story in all its nuances. Copyright was probably unnecessary at the time. There are other examples. Imagine a device which is like a super television—a life size three dimensional image of the performer is brought into your living room. Assume also that performers did not have a property right on their performances and they could not prevent someone from taping nightclub performances, say, of Frank Sinatra and broadcasting it (or selling tapes) to all people with televisions. Small, high quality microphones and cameras would make such copying feasible. Since television would be as good as being at the performance no one would ever go to live performances if watching television was less expensive. The 'efficient' solution, in the sense of Pareto optimality, defined without the inclusion of information and transaction costs, is to provide copies of the performance on television to all possible consumers. This would remove revenue to nightclub performers in the same way that printing, without copyright, removes revenues to authors. Television is not yet a very good substitute for live performances so that the two can coexist even without a special property right. In addition, one cannot usually film a performer without his cooperation and knowledge.

Demsetz has made clear the proposition that the efficient production and consumption of public goods may require exclusion. Pareto optimality is a useless concept for public policy if it is an unreachable goal. The alternative forms of provision of public goods will determine whether exclusion is detrimental to public well-being or not.

A performance right grants copyright to the performer and no one else can make use of that performance on record, tape or film, etc. In 1961 the International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations was adopted. Nineteen countries have joined including the United Kingdom, West Germany, Sweden and Italy. The U.S. and Canada do not have such a right.
The issue boils down to one of property rights. Just as a farmer will not voluntarily cultivate land if any other person can come along and harvest the land, without copyright the author, like the farmer, will not have sufficient pecuniary incentive to engage in the productive act of artistic creation. A copyright is merely a means by which the author is given a property right on his artistic creation. The copyright allows the author to gain financial rewards for his artistic endeavors.

The granting of a property right in 'artistic and intellectual works' is not without its possible costs, however. Since each 'work' is unique, the holder of the property right becomes a monopolist with respect to his 'work'. The work will be published in a manner which maximizes the profits of the copyright holder. Profit maximizing behavior implies that the output of the 'work' will occur at a level of production below that which maximizes the value of the production, causing a deadweight loss. This deadweight loss would disappear if there were free competition in the publishing of this 'work'.

The primary efficiency issue in the economics of intellectual property concerns balancing the incentives to encourage the production of artistic and intellectual works with the cost to society of creating deadweight loss.

---

6As is true for most commodities, property rights over intellectual property are limited. When, for example, the author sells copies of his book, a purchaser may lend it to a friend without needing the permission of the copyright holder. In addition, fair use enables an individual to copy parts of the book without being liable for copyright.

7For the general publishing industry, profit maximization is probably a reasonable assumption. For the writing and publishing of academic journal articles profit maximization does not on the surface appear to be as reasonable. However, our empirical tests are carried out on this latter portion of the industry and the predictions based on profit maximization are borne out empirically. Thus special discussion of not-for-profit institutional behavior seems unnecessary.

The conflict of goals between author and publisher, so common in elementary economics texts, would be easily overcome through negotiation between author and publisher.
The economic justification for granting copyright has been a matter of considerable debate.\textsuperscript{8} The main noneconomic justification for copyright is known as the "moral" right of authors. This justification is often framed in terms of the "just rewards" for one's expenditure of effort. When one examines the asymmetric role of technology in rewarding human efforts, an equity issue which runs parallel to the moral rights literature becomes a compelling argument for copyright. This asymmetry has not been fully realized. For example, in comparing the moral rights of authors to other workers Breyer states:

It is often said that an author should receive the "value" of his work to society--a value that might be measured in terms of what those who benefit from the book would be willing to pay rather than do without it. But few workers receive salaries that approach the total value of what they produce. The social value of the work performed by the man who invents the supermarket, the man who clears the swamp...may be much greater than his pay... We do not feel that owners, managers, or workers in such industries are for this reason morally entitled to higher wages... There is nothing inherently immoral in the fact that many workers are paid less than the social value of what they produce. p. 285

\textsuperscript{8}There is a school of thought which argues that copyright probably grants an unnecessary monopoly: Plant (1934) Hurt and Schuchman (1966), Breyer (1970). These authors argue that being first in the market grants a sufficient monopoly to cover the costs in creating most copyright materials. They believe that the number of additional creations brought forth by copyright would be small and of less value than the unnecessary deadweight loss created in the monopoly sale of those materials which did not need so much protection for their creators to be compensated. However, see Tyerman (1971) for a convincing defense of copyright.

This anti-copyright school can be better thought of as desiring a short copyright life. No one argues with the proposition that effort requires compensation. If monopoly is to be the means for achieving compensation one merely needs to determine the copyright life which maximizes net value to society. If this life is less than the monopoly which exists from being first, then copyright is unnecessary as it is a non-binding constraint, and as well, it causes no welfare losses. Copyright only causes losses if it gives greater duration to monopolies than is optimal. Of course, these losses are not the fault of copyright \textit{per se} but only of the improper life specified in copyright law. If one believes the government incapable of providing proper copyright life, one may decide that copyright should be abolished.
Breyer misstates the purpose of copyright. It is not intended to pay
the author the full value of his work (only perfectly discriminating monopoly
could do this) but instead to give him some payment, where the alternative
is thought to be no payment at all (the monopoly from being first is implicitly
ruled out).

The moral force behind copyright comes about because technology has not
influenced other workers in the same way. The Frank Sinatra example discussed
above indicates one case where a worker is able to receive payment because
technology is not yet advanced enough to disembodify the performance from the man.
Similarly with other workers whose output could become a public good given the
appropriate technology. Academics and talented business executives would be
greatly displeased if a machine could read their minds and provide (business)
advice or scientific advances with no payment made to those individuals. However,
in terms of equity such a result would be acceptable as long as technology
has this effect on everyone. It is the disparate effects of technology that
gives the bite to the equity argument.

Photocopying is a new technology which appears capable of removing payment
to the author. If it can be shown to have such an effect society may wish
to grant stronger protection to copyright holders, for the reasons just
presented. However, we shall see that copyright holders are presently
receiving greater remuneration than is generally understood.

3. **Fair Use - The Main Exception to Copyright Liability**

Fair use is the major exception to infringement of copyright. Does it
benefit one group (researchers, students, reviewers) at the expense of others
(copyright holders) or merely remove the restrictive covenants of copyright
law in those instances where economic well-being of the copyright holder is
not threatened?
A copyright holder's well-being is threatened when his ability to appropriate revenues is reduced. Fair use may do this, for example, if researchers, students, and reviewers made copies of a work instead of buying it. Clearly, in the era prior to electronic reprography, any entire copy would have to be created by hand and the time costs would be very high and probably greater than the cost of purchase. People who hand copy would be likely to copy only small sections. A small portion of the work copied is not likely to be a good substitute for the entire copyrighted work. Allowing fair use in this instance would not likely reduce the revenues of authors by much. 9

In fact, research, reviews and study of intellectual work create interest which may be translated at a later time into the purchase of that (or a different) intellectual work which might not have been purchased without the fair dealing doctrine.

In the present world where modern photocopying machines make it easy to copy an intellectual work, fair use is more likely to decrease the ability of the copyright holder to appropriate revenues. Reviewers and scholars can now copy parts of a work at relatively low costs. Some may copy the entire work instead of buying it. We shall refer to this as the substitution effect. Modern photocopying may incline some individuals not to buy an intellectual work but it is also likely that exposure to various works is increased by cheap photocopying. Individuals who may not

9There is evidence that the courts had this in mind when the concept was constructed. For example, in an important 1841 case Justice Story said: [we] must often, in deciding questions of this sort, look to...the degree in which the use may prejudice the sale or diminish the profits...of the original work. Polsom v. Marsh; 9 Fed. Case 342 (1841).
have been aware of the properties of various intellectual works can sample them by using a copy made on a photcopying machine. This is particularly true among academic researchers using scholarly journals. Many library journals can only be read in the library so that, prior to photcopying, a researcher would read only in journals which he considered most important since taking notes in libraries is a rather unpleasant task and less convenient than reading articles wherever and whenever one wishes. With modern reprographic methods the researcher can make copies of articles in any journal and read them at his leisure (taking notes in the margins of the copy). This increases the number of articles he is likely to read and also increases his exposure to new journals, some of which he may subscribe to eventually. We shall refer to this as the exposure effect.

In addition, it is likely that the price of the intellectual work will reflect the value placed on it by those who buy the work. This value will go up if the purchaser intends to let acquaintances copy parts of the work since he could either charge them in pecuniary terms or by building up goodwill. Also, depositories of books and journals (i.e., libraries) will have their holdings more highly valued by users and should therefore be willing to pay more for the journals (and perhaps increase their total number of subscriptions). The ability of copyright holders to capture revenues from those making copies of their work is of key importance. The negative impact of a reduction in sales will be mitigated by the extent to which they can accomplish this end. We will label this potential force as the addition effect.

Thus the impact of reprography on revenue can be seen as the net impact of the substitution, exposure and addition effects. The substitution effect reduces appropriability, the addition effect increases appropriability and
the exposure effect does not affect appropriability per se, but will influence the well-being of copyright holders. To comprehend this last statement one must realize that the interest in reprography lies in its impact on the transmission mechanism between the use of the intellectual work and the payment to the copyright holder. Both the substitution and price effects act upon this mechanism. The former reduces payments for a given level of use while the latter increases the level of payment. The exposure effect, however affects only the total amount that the good is used, not the transmission mechanism between payment and use. For this reason it is not the proper concern of a copyright policy which was concerned only with appropriability. It is of interest, however, to those concerned with the welfare of copyright holders and society since it influences the resources society spends on intellectual works.

Fair use can be evaluated in terms of these effects. Prior to cheap reprography it is likely that fair use had a small negative substitution effect, a positive exposure effect and a small positive price effect. If this is a proper assessment then fair use had a beneficial impact on copyright holders. Even if fair use had a harmful impact on copyright holders it might still be desirable for the public as a whole. This is because the rapid dissemination of knowledge and research is a valuable goal and such dissemination would be slowed down if students and researchers had to negotiate with the copyright holders before they could make copies. When these transactions costs become greater than the appropiable revenues authors will be unable to collect such revenues. Therefore, no harm would be done and no loss of appropriability would occur in such circumstances.
4. **The Impact of Price Discrimination**

In recent years there has been an increasing tendency toward price discrimination on the part of copyrighted holders of journals. If reprography causes many subscriptions to a journal to be cancelled, the copyright holder is likely to lose revenues unless he can charge a higher price to the subscribers of the photocopied journals. Price discrimination allows copyright holders to capture revenues from users of the journal whether they subscribe or not. If a publisher knows that former subscribers are going to libraries to make photocopies he can raise the subscription price to libraries and recapture some lost subscription revenue. In a world of perfect price discrimination appropriability would be no problem at all. Of course copyright does not allow authors to appropriate consumers surplus except in a very imperfect manner so that imperfect price discrimination may do just as well.

The fact of the matter is that journal publishers are price discriminating such that libraries and institutions pay a higher price than individuals pay for many journals. This will tend to negate the revenue decreasing effects of reprography as well as altering the nature of fair use as a defense against infringement. Fry and White report the average increase in prices charged to institutions and individuals by type of publisher. These results are reproduced below:

<table>
<thead>
<tr>
<th>Table: % Increase in Price From 1969-1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Ind.</td>
</tr>
<tr>
<td>34.8 70.6</td>
</tr>
</tbody>
</table>

These results indicate that the spread between individual and institutional prices grew over the period with most of this growth accounted for by commercial publishers. Additional, and more detailed evidence can be found in the analysis of a sample of leading academic journals. Institution prices were compared to individual prices in 1970 and 1978.\textsuperscript{10} These results are in Table 1. It is readily apparent from this table that the price charged to institutions has gone up relative to that charged to individuals. The t-value for rates (row 7) gives the statistical significance of this difference. The increase in the absolute difference between institutional and individual prices over this nine-year period has been statistically significant at the 99% level of confidence. Looking at the arithmetic difference between these prices may be somewhat misleading due to the inflationary forces prevalent during this period. If inflation impacted on both individual and institution prices to the same extent, the nominal price spread would increase over time. To determine the change in relative prices independent of inflation we formed the ratios of institution to individual prices. These are presented in rows 5 and 6. Observation reveals that these ratios have increased during this period of time. With institution prices 66% higher than individual prices, and institutions making up roughly half the subscribers of many journals, price discrimination clearly has a large effect on revenues.

\textsuperscript{10} Institution prices were taken from the Faxon’s Librarian Guides for each year and the individual prices were taken from the journal on Ulrich’s International Periodicals Directory. List of journals available upon request.
### Table I

**Institution and Individual Prices, 1970-1978**

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>1970 Ind. Price</td>
<td>13.20</td>
</tr>
<tr>
<td>(2)</td>
<td>1970 Inst. Price</td>
<td>16.46</td>
</tr>
<tr>
<td>(3)</td>
<td>1978 Ind. Price</td>
<td>25.17</td>
</tr>
<tr>
<td>(4)</td>
<td>1978 Inst. Price</td>
<td>36.61</td>
</tr>
<tr>
<td>(5)</td>
<td>1970 Ratio Inst./Ind.</td>
<td>1.26</td>
</tr>
<tr>
<td>(6)</td>
<td>1978 Ratio</td>
<td>1.66</td>
</tr>
<tr>
<td>(7)</td>
<td>t-value 1970 price spread</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td>versus 1978 price spread</td>
<td>(99%)</td>
</tr>
<tr>
<td>(8)</td>
<td>t-value 1970 Ratio</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>versus 1978 Ratio</td>
<td>(99%)</td>
</tr>
</tbody>
</table>

N: 54
We cannot be certain as to the causes of this increasing price discrimination but its concurrence with increased photocopying would seem to be more than coincidence. In order to photocopy copyrighted materials one must have access to both a photocopying machine and copyrighted materials. The nexus of these two attributes is found mainly in libraries. As long as copyright holders can identify libraries and charge them higher prices (perhaps varying the price between libraries) they can appropriate revenues from users who switch from subscriptions to libraries.

One difficulty with enforcement of price discrimination is that as the price differential increases, the incentive for libraries to avoid paying the higher price through subterfuge increases. Librarians could order subscriptions under several individual names. The ability of libraries to 'arbitrage' will determine the limit of the market adjustment which can be made by publishers. To prevent such a possibility it would seem prudent to pass legislation making such behavior illegal. Monitoring non-compliance would be inexpensive since publishers or their agents would merely have to inspect the holdings of libraries and determine if they match the subscriptions in the library's name. Some sort of mark could be printed on institutional subscriptions to enhance detection. Penalties merely need be sufficient to deter libraries from such subterfuge.

The present institutional arrangements are advantageous to successful price discrimination. Libraries usually buy their journals through agencies which reduce the costs of ordering or cancelling subscriptions. These agencies can easily be charged the higher price since they are so easy to identify. Only when the reduced transactions costs offered by these agencies are outweighed by the gain to arbitrage will arbitrage become a serious problem.
Another problem concerns the possibility of individuals donating journals as a means of circumventing the price discrimination. However, this could be easily monitored and legislation could be written requiring donors or libraries to pay the rate differential upon donation of the journal.

To the extent that publishers can price discriminate, the category of fair use becomes meaningless. Technically, fair use allows some copying of copyrighted works, under given conditions, to occur with no payment made to the copyright owner. Under a regime of price discrimination, the copyright owner is being paid for copying which comes under the rubric of fair use. It must be concluded that for journals which price discriminate, fair use is not a meaningful concept. When price discrimination is used to generate payments for use of copyrighted material, no distinction is made between users who copy under fair use guidelines and those who don't.

It would be wrong, however, to conclude that price discrimination has perverted the intent of fair use. Action to resurrect fair use as a viable defense would indicate a lack of understanding of the economic function of the fair use doctrine. It is not important that fair users not have to pay copyright so much as it is important that they not be hindered in their research by negotiated copyright payments. Price discrimination is a collection method which does not require lengthy contractual arrangements between users and copyright holders. The fair user is probably completely unaware that he is paying for his copying activities. An arrangement where producers of intellectual property are reimbursed and consumers are not hindered in their use of such materials, is one which should be encouraged.
Another impact of price discrimination is that borrowers of copyrighted materials who do not make copies pay for their use of the materials just as copiers do. This payment can be considered a distortion of the decision not to grant a public lending right to authors. Discriminatory pricing makes such a lending right redundant.

5. A Copyright Clearance Center

An alternative to price discrimination would be strict adherence to the copyright law under unitary pricing. There are two major stumbling blocks to strict adherence. The first is the high negotiating costs relative to the value of making a copy. People making photocopies want them immediately and are not prepared to wait several weeks to get the copyright holder's permission, even if the royalty payment itself is not objectionable. The second is the monitoring costs involved in trying to prevent people from copying the intellectual property without any payment (theft).

In hearings prior to the 1976 copyright law, the U.S. congress made clear that it wished to have an organization in place to simplify the copyright negotiations. In response to this a non-profit Copyright Clearance Center was recently set up. From the 'Summary':

Organized principally by the efforts of non-profit and professional society publishers with help from a few large photocopy user organizations, CCC was initially funded by gift monies received from publishing organizations of all types. Nearly half were non-profit organizations seeking to strengthen the principles of copyright and to help establish a system to uphold their publication income.

---

11 Waite, David, P. "A Summary of the Copyright Clearance Center's First Year Operating Experience," p. 10.
Publishers register titles with the clearinghouse. Lists of these titles, along with copyright payment schedules, are sent to user organizations (libraries, corporations, government agencies and information services). User organizations voluntarily register with the CCC (Copyright Clearance Center) and report on their photocopying practices, making the appropriate payments. The CCC keeps 25 cents per reported photocopy as payment for its collection services. The important aspects of this system are (1) voluntary payments and (2) payment fees determined by publishers.

The largest costs associated with strict liability are monitoring of users by copyright owners. The CCC, by abrogating this responsibility, avoids the serious problems associated with effective policing of users. Such a policy engenders other problems, particularly the non-reporting of photocopying activity by users. Such a policy also rules out the single greatest advantage of a centralized agency i.e. the elimination of the need for duplication in monitoring costs which would be created if many publishers independently tried to monitor users. However, some costs might be saved if there were economies of scale in processing forms, collecting payments and disbursing these payments.

The performance of the CCC, to date, has not proven the workability of the clearinghouse concept for increasing publisher profits or society's welfare. Over 3800 publishers were solicited to register their serial

---

12 In a paper purporting to analyze the imposition of copyright or user fees upon the population of library users, Ordover and Willig conclude that some positive fee must be welfare improving and profit maximizing. This result is based on a model with very restrictive assumptions. The model also neglects the influence of transacting cost, probably the single most important determinant of the impact of user fees. Because of these difficulties the results of Ordover and Willig cannot be taken as a serious challenge to the findings of this paper. See: Ordover and Willig [1978] and Liebowitz and Margolis [1980].
publications and yet after twenty-one months of operation only 335 publishers had enrolled in the system; 923 user organizations are currently registered. The CCC, at last report, received payment for an average of 14,750 copies per month. The latest "projected" break-even point is estimated to require 85,000 copies per month; in other words, almost a six-fold increase. The CCC has made up this deficit by generating contributions from various organizations and withholding some of the royalty payments to publishers (effectively charging them 50 cents per copy). The average fee paid by users per reported photocopy was $1.53. CCC's paid staff totals three people.

The projected break-even point reported by the CCC is somewhat suspect. In June 1979 the CCC claimed to be approaching 50% of the self-sustaining volume of photocopies. The November 1979 report claimed to be at the 17% level, invalidating the earlier projections. Even if the CCC manages some day to be able to cover its transactions costs its worth to society will be unproven since the costs to users of filling in forms and monitoring their own usage is likely to be much larger than direct CCC costs and this cost is not included in the above calculations.

The CCC believes its poor performance to date is due to the fact that certain users are not accurately reporting their photocopy totals, hardly surprising given the nature of enforcement. Only 30% of the registered users reported any photocopy activity over the entire period. Seventy-one percent of users reporting activity were in private industry. Academic libraries, on the other hand, only 15% of whom had registered with the CCC, had only 7% of those registered reporting photocopy activity.¹³

---

¹³One reason for this very low degree of reported photocopying is that much of the photocopying activity is considered to be fair use by libraries in these institutions.
It should be pointed out that a centralized clearinghouse could be set up for a somewhat different and more nefarious purpose. It could be the case that a clearinghouse might not allow individual publishers to determine the copying fees for their works but would instead take over this responsibility itself. Allowing such control would be tantamount to allowing the reduction or elimination of competition between copyright holders and a large increase in the monopoly power and deadweight losses due to the copyright system.\(^\text{14}\)

Thus the CCC must be viewed as an economic failure since it appears unable even to pay for itself and thus must fail in its attempts to compensate copyright owners. Projected future viability of the CCC must be viewed with skepticism given the past performance of these projections and their underlying assumptions. This conclusion is enhanced for those countries which are small relative to the U.S. A country such as Canada, one-tenth the size of the U.S., could not hope for a self-supporting CCC.

6. **The Institutional Choices**

We have so far examined two alternative but not mutually incompatible institutional arrangements for the enrichment of copyright owners. There are many other suggested possibilities for collecting revenues such as a tax on photocopy machines or granting a compulsory license to libraries or other user groups. We cannot attempt to examine all possible alternatives. In fact, recent changes in technology allow sufficient prescience to conclude

---

\(^\text{14}\) This problem has been recognized in regards to performing rights societies such as ASCAP and BMI. See Broadcast Music, Inc. v. CBS, 441 U.S. 1 (1979).
that the institutional arrangements presently under analysis will be short-lived. The use of computers and the future abundance of home terminals will alter the parameters affecting the production function of using intellectual properties and collecting revenues for this use. Journals may disappear as physical entities to be replaced by computer information retrieval systems with, as seems likely, an editorial board serving functions similar to those performed today. Negotiations between users and copyright holders will still be costly and so one would expect a flat fee charged to all users of a given class. Billing will occur immediately and at very low cost. Fees for use as well as fees for making copies are quite likely possible. Concepts such as fair use or public lending rights would likely be unnecessary.

However, until the future arrives we must use our present institutions. I have argued in the last section that the CCC must be viewed as an economic failure. From this one can conclude that the CCC should be dropped as a possible institutional arrangement. Of course other arrangements are not ruled out. However, the existence of price discrimination alters the relationship between revenues to copyright holders and use of copyrighted materials in a manner beneficial to copyright holders. The increase in price discrimination concurrently with increased photocopying suggests a causative relationship. If this is so the urgency in enhancing copyright protection is certainly diminished, if not negated.

Price discrimination is the market's response to changing conditions. The increased revenues from price discrimination dwarf any projected revenues coming from the CCC. It is possible that some government imposed institution might

15 Though it might be a political success since both copyright holders and libraries, the two special interest groups most affected, appear to be somewhat appeased at present.
have greater benefits. However the institution which has been chosen has not. Is there any reason to believe that a future governmental choice would be better? Would not a more fruitful approach be to enhance the ability of the market to adjust to changing conditions, as proposed in section 4, instead of trying to patch up what in the copyright holder's incorrect view is a leaky system?

7. Conclusions

We have endeavored in this paper to demonstrate that the debate over special copyright protection against photocopying has neglected the important market adjustment of price discrimination. The emergence of price discrimination has benefitted copyright holders and altered the impact of current copyright law. The justifications for enhancing copyright protection are weakened by price discrimination. The particular institutional attempt to enhance copyright protection is shown to be economically inefficient. We recommend that the government turn its efforts to enhancing the market's ability to react, instead of creating artificial markets.
References


Waite, David P. "A Summary of the Copyright Clearance Center's First Years Operating Experience." Copyright Clearance Center Report.
COPYRIGHT AND COMPUTER SOFTWARE

by

John P. Palmer

The Centre for Economic Analysis of Property Rights
The University of Western Ontario
London Canada

March 1983

Helpful comments were provided by Albert Link and Stephen Margolis.
The invention of and the growth of the popularity of sound recordings markedly altered the ways in which composers were and were not compensated for their creative efforts. Nevertheless, a study of the sound recording industry in 1915 would have found it to be rapidly growing and developing. Arguments against copyright in the industry would have pointed to this growth, questioning the need for more explicit legal protection for composers and record producers. But a market failure (in the sense of narrowly and imprecisely defined property rights) did exist even though the industry was so vibrant. Copyright protection was extended to cover this new medium for the expressions of ideas to correct the market failure.

It is time for Parliament once again to consider seriously the extension of copyright protection to cover new media for the expressions of ideas to correct new market failures in the computer software industry. In doing so, it must be careful to include protection for all software, including firmware (the embodiment of a computer program on a silicon chip) and it must also take care to extend only those provisions of the Copyright Act to software for which the net social benefits would be positive.

Using a basic supply and demand approach, this study calculates that the social benefits of providing copyright protection for traditional software would be at a minimum approximately $40,000 per year. Subtracting $20,000 per year for administrative costs leaves net benefits of $20,000 per year as the minimum net social benefits.

Using the same procedure, the net benefits could be as high as nearly $2 million per year. At least with the data available it would not
be easy to argue against either of these estimates. Each of them is extreme, and it is more likely that the true net gain lies somewhere between these two extreme figures. This study sets out the basis for these estimates, and discusses important qualifications of them.

As Demsetz has so clearly emphasized in his scathing criticism of Arrow, discussions of policy alternatives must keep in mind feasible standards for comparison. In this regard, this study will not be comparing each alternative with some idealized, perfectly competitive norm; rather it will begin by comparing alternatives with the status quo, which is based primarily on the use of market, technological, and trade secret protection. As the discussion progresses, it will include additional alternatives in the comparisons.

A. Need for Additional Protection?

The computer software industry has been growing dramatically during the past decade. It is reasonable to ask whether there is any need for the protection of software beyond the protection currently available technologically and through rapid marketing or trade secret protection. If the industry has grown rapidly and offered sufficient incentives for entry and technological development, it has sometimes been argued, then there is no justifiable basis for creating even a limited property right and monopoly in specific computer software. The crux of this argument is that providing intellectual property protection in addition to what is available would provide no additional benefits for society, and that allowing even a limited monopoly right would generate some extra costs for society. Of course, one relevant consideration is the counter-factual question: how fast might the industry have developed had more legal protection been available for software?
In an exploration of the need for intellectual property protection of computer software, Miller surveyed software producers to determine what they believed to be effective means of protection. The firms were asked to indicate the degree of effectiveness of alternative means of protecting their software. Their responses are indicated in Table 1. As can readily be seen, most firms found a lease with a confidential disclosure clause or trade secret licencing to be the most effective modes of legal protection. Interestingly 58% of the respondents claimed to have tried copyright protection yet there had been fewer than 1000 copyright registrations of computer software in the United States when the survey was conducted. It seems a bit peculiar that respondents would claim that copyright protection had any effectiveness at all in the U.S. prior to 1978 in view of the fact that the U.S. copyright law did not clearly include software and that the only legal case involving copyright and software did not directly support the copyrightability of software. One possibility is that software producers found that copyright protection of their instruction manuals provided some protection for their software. Another possibility is that the respondents expected that copyright protection, if available and if supported by the courts, would have some modicum of effectiveness of protection.

In a later study carried out for the Commission on New Technological Uses of Copyright (CONTU), Miller surveyed a much larger sample of software producers using a somewhat more extensive questionnaire. As with the 1973 survey, the 1977 survey was accompanied by a letter explaining that the major interest in the survey concerned the possible effects of copyright protection. The results, presented in Table 2, indicate decreased use of copyright between 1973 and 1977 (compare the last column of Tables 1 and 2) as the number of firms
### Table 1
Preferred Modes of Legal Protection (Figures indicate percentage of respondents answering in each category)

<table>
<thead>
<tr>
<th>Mode of Protection</th>
<th>Not At All Effective</th>
<th>Somewhat Effective</th>
<th>Very Effective</th>
<th>Completely Effective</th>
<th>Not Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lease with a Confidential Disclosure</td>
<td>.03</td>
<td>.23</td>
<td>.35</td>
<td>.16</td>
<td>.23</td>
</tr>
<tr>
<td>Trade Secret License</td>
<td>.13</td>
<td>.16</td>
<td>.26</td>
<td>.10</td>
<td>.35</td>
</tr>
<tr>
<td>Copyright</td>
<td>.09</td>
<td>.26</td>
<td>.16</td>
<td>.07</td>
<td>.42</td>
</tr>
<tr>
<td>Physically Limiting Access to Technology</td>
<td>.07</td>
<td>.16</td>
<td>.20</td>
<td>.13</td>
<td>.44</td>
</tr>
<tr>
<td>Cryptographic Coding</td>
<td>.13</td>
<td>.10</td>
<td>.07</td>
<td>0</td>
<td>.70</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software Lock</td>
<td>0</td>
<td>0</td>
<td>.08</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td>Controlled Support</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td>Patent</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.97</td>
</tr>
</tbody>
</table>

Source: Miller, n. 2 infra.

### Table 2
PREFERRED MODE OF PROTECTION
(Figures Indicate Percentage of Respondents Answering in Each Category)

<table>
<thead>
<tr>
<th>Mode of Protection</th>
<th>Not at all Effective</th>
<th>Rarely Effective</th>
<th>Somewhat Effective</th>
<th>Fairly Effective</th>
<th>Very Effective</th>
<th>Completely Effective</th>
<th>Frequency of Use*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent</td>
<td>.82</td>
<td>0</td>
<td>0</td>
<td>.18</td>
<td>0</td>
<td>0</td>
<td>.04</td>
</tr>
<tr>
<td>Copyright</td>
<td>.55</td>
<td>0</td>
<td>0.05</td>
<td>.1</td>
<td>.15</td>
<td>.15</td>
<td>.2</td>
</tr>
<tr>
<td>Trade Secret</td>
<td>.29</td>
<td>.05</td>
<td>.14</td>
<td>.24</td>
<td>.14</td>
<td>.14</td>
<td>.21</td>
</tr>
<tr>
<td>Release of Object Program Only</td>
<td>.17</td>
<td>0</td>
<td>.04</td>
<td>.08</td>
<td>.33</td>
<td>.38</td>
<td>.3</td>
</tr>
<tr>
<td>Know-How Requirement</td>
<td>.28</td>
<td>.17</td>
<td>0</td>
<td>.17</td>
<td>.17</td>
<td>.22</td>
<td>.13</td>
</tr>
<tr>
<td>Cryptographic Coding</td>
<td>.5</td>
<td>0</td>
<td>.17</td>
<td>.25</td>
<td>.08</td>
<td>0</td>
<td>.4</td>
</tr>
<tr>
<td>Other Means of Limiting Access</td>
<td>.27</td>
<td>0</td>
<td>0</td>
<td>.13</td>
<td>.06</td>
<td>.53</td>
<td>.17</td>
</tr>
</tbody>
</table>

*The figures in this column relate to the entire sample.

Source: Miller, n. 3.
surveyed quadrupled. Interestingly, of those firms which used copyright protection, the views concerning its effectiveness became more polarized with many more finding it ineffective and, as well, more finding it to be effective. A plausible explanation of this result is that as firms attempted to protect their software with copyright, some were quite disillusioned because the copyright protection was ineffective against imitators' use of an algorithm. The less sophisticated firms may have been attempting to use copyright to protect their algorithms and been disappointed; the more sophisticated firms have made extensive use of trade secrecy tactics (e.g., encryption and release of only the object program) to protect their algorithms, and they know to use copyright to protect only a particular expression of an algorithm. Miller's explanation of these results is similar to the one presented above: "We explain these developments by observing that as copyright usage has increased, more respondents have used this relatively inexpensive, accessible mechanism to try to protect more programs that are easily designed around."4

Morgan's comment on Miller's result is:

This conclusion should have come as no surprise to anyone in this country, since what it means is that:

(i) in practice, infringement of title is impossible to detect or stop (e.g. a software house adapts programs written for client A for use by client B); or
(ii) title can be protected physically or technologically (e.g. delivery to the licensee only, of an object code version of the program suitable for running on his machine, and withholding all source language versions); or
(iii) title can be protected by contract.5

In other words, Morgan is asserting that copyright protection offers firms little, if anything, more than is presently available, i.e., it offers no private benefits and hence no incentive for the private sector to generate any social benefits.
There are two reactions to these survey results and Morgan's interpretation of them. The first is that if copyright protection of software offers no private benefits then it follows that it also imposes little or no social costs (other than legal and administrative costs). The private benefits of the monopoly due to copyright exist only if the monopoly power can be and is exercised. If these private benefits are small, then the deadweight losses due to the lost consumers' surpluses are likely also to be small, and concern about the creation of monopoly distortions is a gross exaggeration. This is an important point deserving reiteration. Critics of intellectual property protection of computer software (or anything else, for that matter) cannot have it both ways. They cannot argue it has no benefits and yet would impose immense social costs through resource misallocation. The only way legal intellectual property protection can have no social benefits is if it creates no incentives for people to generate intellectual output, i.e., if it creates no private benefits. And if it creates no private benefits and no incentives for resource reallocation, it cannot then be criticized for creating serious resource misallocation. 6

The second reaction to Morgan's remarks is to question the validity of his (implied) conclusion that the social benefits of copyright or similar protection for computer software would be zero. Reading further in Miller's first study, one sees the following statement:

Eighty-seven percent of all respondents could not think of a single instance in which computer programs representing a significant level of innovation were not developed or marketed because of inadequate protection. The companies that thought the law had been a barrier cited examples in which fear of easy plagiarism or unauthorized disclosure might prevent recoupment of development costs. The situations cited involved such techniques as paging programs for virtual memory computers, an innovative approach to developing multiprogramming capability on the IBM 360/20, and systems software for organizing computer program libraries. 7
Despite whatever biases may have existed in his survey due to its small size and its use of only ADAPSO members, Miller's results indicate that a non-negligible number of firms were aware of instances in which technological development had been impeded for fear that the lack of intellectual property protection might give rise to misappropriation.

In his 1977 study, Miller found that, "seventy-four percent of the sample had never rejected or abandoned a program because of the presence or absence of protection and 65 percent would not change their marketing even if protection were provided."[^8] "Seventy-seven percent of the sample knew of no instance of aborted marketing or development."[^9] It appears from the results of Miller's two studies that an increasing number of instances was occurring in which computer software was not developed or marketed as a result of the lack of protection. It is unfortunate that Miller did not specify in this question what he meant by legal protection, but as was indicated earlier, his cover letters, along with the other questions in the surveys indicate that the surveys were being carried out to study copyright protection, and one assumes that most firms answering these questions would have interpreted legal protection as meaning copyright protection.

One of the reasons that fewer firms were unaware of impeded software development and marketing was that proprietary software packages were just becoming increasingly important in the software industry during the mid-1970s. The projection by Business Week that proprietary software packages will take on an even greater role in the industry suggests that the lack of legal protection for these packages will have an even greater effect of impeding development in the future (see Figure 1).
Figure 1

How sales of standard packages will overtake custom software

Source: Business Week, September 1, 1980, p. 56.

Table 3

SIGNIFICANCE OF SOFTWARE PROTECTION BY FUNCTION*

<table>
<thead>
<tr>
<th>Function Description</th>
<th>0-10</th>
<th>11-25</th>
<th>26-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) General business and financial applications (e.g., accounting, inventory control, payroll)</td>
<td>.17</td>
<td>.33</td>
<td>5</td>
</tr>
<tr>
<td>(b) Business planning operations (e.g., planning models, simulations, operations research)</td>
<td>.5</td>
<td>.24</td>
<td>.26</td>
</tr>
<tr>
<td>(c) Complex production/distribution control operations (e.g., linear programming)</td>
<td>.55</td>
<td>.19</td>
<td>.26</td>
</tr>
<tr>
<td>(d) Engineering and scientific applications</td>
<td>.53</td>
<td>.19</td>
<td>.28</td>
</tr>
<tr>
<td>(e) Data and statistical analysis</td>
<td>.51</td>
<td>.17</td>
<td>.31</td>
</tr>
<tr>
<td>(f) Project management and control</td>
<td>.54</td>
<td>.23</td>
<td>.23</td>
</tr>
<tr>
<td>(g) Systems software (e.g., compilers, monitors, new techniques for more efficient machine utilization)</td>
<td>.4</td>
<td>.17</td>
<td>.43</td>
</tr>
</tbody>
</table>

*Figures indicate percentage of respondents answering in each category.

Source: Miller, n. 3, p. 22.
Miller's more detailed results support this projection that copyright will become more important in the future. He segmented his sample according to the type of software produced by each firm and asked the firms to indicate the significance of protection for their product(s). His results are presented in Table 3. Miller summarizes the results by saying,

"With allowances for a difference in the way in which the data is expressed, the CONTU results are strikingly similar to those of the 1973...survey. What they show is that the more universal and widely marketed the program the more important is protection. This is a characteristic of general business programs, which can be addressed to such functions as payroll and receivables anywhere, and also of systems software in which a program can be used for a particular computer in a variety of installations. On the other hand, the more technical and unique the program the less significant protection appears to be. This finding is consistent with information which was supplied to us about the programs which were being marketed."[10]

And in response to an open-ended question, "The comment most often repeated was that an apparent lack of interest in legal protection was related to the fact that they did not market proprietary software."[11]

In conclusion, the software industry has grown rapidly in the past decade, even in the absence of explicit legal protection for its intellectual property. An important question for policy prescription, however, is how fast would the industry have grown had better protection been available? If the industry could have grown even faster with intellectual property protection, there is a reasonable chance that protection would have had some net social benefits. A related question addresses the future trends in the industry. We have seen that industry experts expect that there will be even more emphasis in the future on proprietary software packages, precisely those which Miller found to be more amenable to explicit legal protection because these programs require wide-scale marketing and are hence more costly to protect using trade secrets. What type of legal protection will be the most valuable to society for firms in this segment of the industry is vague in Miller's studies, although
it is reasonable to assume that Miller and his respondents had copyright protection in mind. The indication from Miller’s studies and others is that the continued growth of the proprietary software package segment of the computer software industry has made and will continue to make the potential net social benefits from legal, possibly copyright, protection of computer software increase over time. 12

B. Framework for Analysis

Before examining in greater detail net social benefits of various types of legal protection for computer software, in this section a theoretical structure for analyzing the net benefits is set out: Assume that there is a downward-sloping demand curve for software. Assume also that the long-run supply curve for computer software is horizontal and that software is produced under competitive conditions. Finally, assume that initially the only protection available for software producers is trade secrecy or technological protection.

One of the reasons that intellectual property protection can be beneficial to society is that it reduces the average costs of producing computer software by reducing the costs of protecting against misappropriation by imitators. As we have seen, many computer programs would continue to be produced, possibly without much change in the costs of production and marketing, were intellectual property protection available. Nevertheless the major reason that programs were not produced which might have been produced under explicit legal protection was that the costs of private protection against misappropriation were expected to be greater than the expected gains from producing the software. Furthermore some programs currently being produced and marketed could be produced and marketed at lower costs were legal protection available. 13 As a result, on average the provision of more explicit legal intellectual property protection would reduce the average costs of producing software, and in Figure 2 would shift
the supply curve downward to $S'$. This reduction in costs would generate additional social benefits (ignoring for the moment legal and administrative costs) equal to the area of the shaded trapezoid in Figure 2, $P^*GFP'$. This area, $B$, can be shown to equal

$$B = \frac{\partial Q}{\partial E} + \frac{(\partial Q)^2 P^*}{2Q^*E},$$

where $\Delta P = P^* - P'$, $\Delta Q = Q' - Q^*$, and $E = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{\Delta P}$.

Examination of equation (1) reveals that if $P^*$ and $Q$ are understated, and if $E$ and $Q^*$ are overstated, the resulting estimates of the social benefits of protection will be biased downward. This is the procedure to be followed in the study because estimate of the values of each of the variables is quite imprecise. If we stack the deck against finding significant social benefits and still find some, confidence in the existence of these benefits is increased. Later some possible maximum values for $B$ are calculated by biasing the estimate upwards.

Before this framework is used to analyze policy options, however, some of the potential problems with using it should be noted:

1. **Non-Horizontal Supply Curves**

Even in the long run the supply curve for computer software may not necessarily be horizontal. If there are some inputs for the production of software which would be scarce even in the long run, even allowing for factor mobility, then the costs of producing computer software would increase as a result of entry and attempts within the industry to increase output. In this case, the supply curve would be upward sloping, and the social benefits of any given cost reduction for developing and marketing software could be smaller than those estimated using equation (1).
2. Demand Curve Shifts

Estimating social benefits using equation (1) assumes that intellectual property protection would have no effect on the demand for computer software. This assumption may not be an accurate reflection of reality. Intellectual property protection could increase the demand for computer software in three ways. The first is that additional legal protection would increase the black market price of pirated programs, thus increasing the demand for the protected programs. The second is that to the extent that intellectual property protection of computer software reduces search costs, the area of the trapezoid will be reduced, and equation (1) will understate the social benefits of the protection. Finally, if intellectual property protection increases the average quality of software, it will also on average increase the demand for software and could increase social benefits in this way as well, depending on the costs of producing higher quality software.

C. Patent Protection

Despite general U.S. Supreme Court holdings that specific software is unpatentable, there is continued and considerable interest in the patentability of software. One reason for this interest is that the court has continually denied any implication from its decisions that all software is unpatentable; another is that recent decisions have indicated that in limited situations, software is patentable. In Canada, the Patent Appeal Board now holds software to be unpatentable, yet this holding has been appealed to the federal courts and numerous other cases before the patent appeal board are pending. Evidence of continued interest in the patentability of software comes from the fact that by 1977 only 1205 computer programs had been registered for copyright protection in the U.S. At the same time, Gemignani estimates that approximately 450 patent applications for computer software are filed each year. And a
recent article by Novick and Wallenstein argues strenuously for the patentability of algorithms.17

The major reason for this continued interest in patent protection for computer software is that it would allow for considerable private benefits to software developers by allowing them monopoly power over an algorithm, regardless of how that algorithm is expressed. The expected monopoly gains from such protection would be considerably greater than for copyright protection, which grants a limited monopoly over only the unique expression of an algorithm and which does not afford any protection to the algorithm itself. Because the potential monopoly gains are great, developers are willing to devote considerable resources to attempting to achieve them— even in the face of a low probability of success. Unfortunately for society, most of these resources devoted to seeking patented software are currently legal and political. If they were technological, increasing research efforts toward finding new, patentable software, the arguments in favour of software patentability would be stronger in that such resources would aid in promoting economic growth, the primary rationale for intellectual property protection. If patent protection for computer software were clearly available, of course, resources would be shifted from the legal and political arenas to the research and development arena, and technological growth might well be fostered. Nevertheless, patent protection might just as easily impede economic growth by allowing the preemption of mathematical and scientific principles and by defeating one of the purposes of patent protection— the dissemination and wide-spread availability of ideas. Furthermore the legal costs of deciding who had prior claim to an algorithm seem, ex ante, to be very high. Finally, in considering the possible patentability of computer software, other social costs must also be considered. The courts have consistently
expressed a fear that granting such patents would preempt a theorem or principle not only as it is applied in the software industry, but for any application.

1. Administrative Costs

In 1979 the Canadian Patent Office received approximately 23000 patent applications and had a budget of between eight and nine million dollars, meaning that the average cost to taxpayers was approximately $350 per application. If anything, the incremental costs per application for computer software would be greater than this average figure. Computer programs can be extremely long and complex, necessitating long hours of skilled examination and comparison with prior art by experts before deciding whether or not they would merit patent protection. The current delay from application to decision averages slightly more than two years, and one can imagine that for complex software this lag would increase. These government costs do not represent all of the social administrative costs of patents, however. In addition, firms spend, on the average, more than $1500 per application for legal fees simply to file a patent application. Additional costs would be imposed on society by the lack of institutional competence within the judicial system to decide between competing arguments concerning alleged infringement of complex and lengthy algorithms. In comparison with technological or trade secret protection of algorithms, these administrative costs may nevertheless be low: trade secret protection requires that reasonable steps be taken to safeguard a secret, and these reasonable steps could easily amount to more than two or three thousand dollars per program; also, technological protection requires the development of additional coding or technology, which could also utilize resources amounting
to more than either the average or the incremental costs of patent protection. In fact, it is this very saving in administrative costs of patent protection over trade secrecy and technological protection which adds to its desirability in its traditional coverage in other industries.

In conclusion, patent protection for computer software has some advantages over trade secret and technological protection. Its administrative costs may be lower and it may encourage broader dissemination of ideas. It is not at all clear, however, that the extended protection offered by patent over trade secret or copyright protection is socially desirable because it might allow the preemption of scientific or mathematical principles. Although trade secrecy also grants a type of monopoly over basic principles, it does so only so long as the secret is kept and undiscovered. Trade secrecy does not preempt independent discovery, and as a result allows for considerable incentive to utilize basic principles without the fear of an infringement suit.

D. **Copyright Protection**

The Copyright Act does not generally protect the ideas expressed in a creative work; it protects only the unique expression of those ideas. For computer software, this distinction between ideas expressed and the expression of ideas means that algorithms would not be protected by copyright. Immitators would be free to use an algorithm so long as they developed their own expression of it and did not slavishly copy its expression from its originator. Copyright protection of software would be a very limited form of protection. To the extent that the commercial value of software lies in its ideas expressed rather than its expression of ideas, copyright would offer no private or social benefits, and trade secrecy would remain the preferred mode of protection.
Contrary to patent or industrial design protection copyright protection of computer software would not require that the interpretations of and rationales for the act be substantially altered. Basically all that would be required is an extension of the definitions of fixation and infringing copies to include new media for the expression of ideas. These definitions need only state that fixation or unauthorized copying would take place even if the media involved included new technology such as magnetic tape cartridges or read-only memories in silicon chip microprocessing units regardless of whether the fixation or copy ever was in a form intelligible to the unaided human eye or ear and regardless of whether it had the potential of generating fleeting images on a video display unit or a printed copy.

1. The Benefits of Copyright

In order to calculate a value for $B$ in equation (1), we need values for $P^*$, $Q^*$, $E$, and $Q$. Each of these variables is discussed in turn.

a. The Average Price of Computer Software Packages

International Computer Programs, Inc., conducted a survey in 1978, the results of which indicated that about 41% of the software products sold for less than $10,000; 66% for less than $25,000; and 85% for less than $50,000. Interpolation between the numbers reported yields a median price near $15,000 in 1978. The distribution of price is undoubtedly skewed, more heavily represented by lower-priced software, and the median is consequently also below the mean. Using $15,000 for $P^*$ in 1978, as a result, is undoubtedly too low, biasing the results against finding sizeable social benefits.
In carrying out the calculations, Statistics Canada data reported for the six years 1972, 1973, and 1975-78 were used. Certainly $P^*$ increased during that time period as a result of inflation. Without any strong justification other than a desire to allow for inflationary forces, let $P^*$ equal $10,000 in 1972, and increase by $1000 for each of the years for which other data are available. Overall, for the entire time period, let $P^*$ equal $12,000. Despite the lack of hard evidence for this figure, it is plausibly low in light of the data available and should be adequate for calculating a downward-biased value for $B$.

b. The Number of Transactions in Computer Software Packages

Statistics Canada has reported the total sales revenue from software packages for the years 1972, 1973, and 1975-78. Dividing the revenues for each of these years by the average price of software for the respective year will yield a value for $Q^*$ in each year. The total for all of these years comes to 6,770. This figure represents the number of transactions, not the number of different programs. It includes each sale, even if one package was sold ten times. That is what should be included since the framework for analysis was constructed on the basis of Miller's results that it is particularly among multiple sales software packages that costs of development and marketing can be reduced with the provision of copyright protection. Note, too, that to the extent that the estimate of $P^*$ is biased downward, the estimate of $Q^*$ will be biased upward, imparting a further downward bias to the estimate of $B$. 


c. Elasticity of Demand

There are no estimates available of the price elasticity of demand for computer software. There are, however, numerous estimates of price elasticities of demand available for other products, most of which are well below five in absolute value. In the initial calculations we will use \( E = 10.0 \), thus likely biasing downward the estimate of \( B \).

d. Increased Number of Transactions

In Section A of this paper, Richard Miller's studies were quoted extensively, and they indicated that the percentage of firms which knew of no software which had not been developed and marketed as a result of the lack of protection had fallen from 87% in 1973 to 74% in 1977. Based on these findings, the higher figure is used (to bias downward \( \Delta Q \) and hence \( B \)), and it is assumed that 13% of the firms knew of at least one instance in which the lack of copyright protection proved to be a barrier to the production of a software package. Transferring this percentage to Canada yields an estimate that approximately 90 firms knew of instances in which software was not developed and marketed due to the lack of copyright protection. One should not, however, use 90 as a value for \( \Delta Q \). On one hand it is unreasonable to expect that in each of these instances the program would have reached the market even if copyright protection had been available. On the other hand, some of the firms presumably knew of more than one instance. And finally, \( \Delta Q \) should be a measure of foregone transactions, not foregone new programs as a result of the lack of copyright protection. To get an idea of the number of foregone transactions, assume that only 50 new programs would have been produced and marketed had copyright protection been available. Further assume that
each of these would on average have been sold four times, yielding a value for \( \Delta Q \) of 200. This figure is quite likely an understatement of the true \( \Delta Q \) in that it is calculated on the basis of small estimates of foregone transactions involving new programs, and it completely disregards foregone transactions of existing programs.

Letting \( Q^* = 6770 \), \( P^* = $12000 \), \( \Delta Q = 200 \), and \( E = 10 \) means that, from the formula for elasticity, we are also letting \( \Delta P \) equal about \$35. In other words, the values for these variables imply that if copyright protection had been available, the cost saving and hence the price reduction would have been on average only \$35, a very small figure for programs costing on average \$12,000. This cost reduction seems particularly small in the light of earlier-mentioned anecdotal evidence that trade secrecy can sometimes add several thousand dollars to the cost of a program. In terms of Figure 2, \( \Delta P \) is the vertical height of the trapezoid, \( B \), which we wish to measure. Finding such a small value for \( \Delta P \) is further evidence of our success in biasing downward the calculation of \( B \), the social benefits of copyright.

Using the figures discussed above the lost social value from not having had this protection is calculated as slightly more than \$240,000 for the six-year period, or, roughly \$40,000 per year. If one assumes instead that \( E = 5 \) or that \( \Delta Q = 400 \), these estimates would double to over \$80,000 per year.

Suppose instead that one wished to establish that the expected benefits of extending copyright protection to computer software would be small (this in fact was the initial position of the author). Using the same data source, one would then attempt to select values for each of the variables which would bias upward the value of \( B \) and show that it was still small, i.e., one would attempt to show that the upper bound is low. In that regard, one might not view \( P^* = $20,000 \) and \( \Delta Q = 1000 \) as unreasonably high, nor \( Q^* = 5000 \) and \( E = 2 \) as
unreasonably low. These figures, however, generate a value for B of $11 million over six years or nearly 2 million dollars per year. This figure is probably as much of an overestimate as $40,000 per year is an underestimate of the social benefits. While the probability distribution of all the possibilities within this range is unknown, so one cannot with confidence settle on any single number, one can, however, assert with confidence that the benefits are very likely to be greater than $40,000 per year (subject to later qualifications) and less than $2 million per year.

2. Administrative Costs of Copyright

In the fiscal year 1979-80, the Canadian Copyright Office registered 7521 works and had a total budget of $336,000, yielding an average of only $45 per registered work. Recalling that Canadian law does not require registration as a precondition for copyright protection, one can readily see that the average government cost per work protected by copyright is far less than even $45, as compared with over $350 per patent application. A further consideration is that if each new program were sold only, on average, four times, then the average government cost for each new transaction would be most certainly less than $12 per transaction.

Because registration is not likely to be required in Canadian copyright law, it is inconceivable that copyright protection of software would require more than $20,000 of government expenditure. This additional amount might possibly arise if computer software developers sought registration (to establish the presumption of originality) in such magnitudes that the Copyright Office found it necessary to hire an additional clerk and to provide additional filing capabilities. At an average registration cost of $50 per program, it would
take registration of 400 programs per year to generate additional costs of $20,000 per year. Even in the United States, which generates substantially more software than Canada and which requires registration for copyright protection, the most programs registered for copyright protection in any year was 282, in 1972/73. 25

There are other potential social costs of copyright in addition to the government's administrative costs. The time costs for copyright protection, however, are negligible; the protection begins from the moment of fixation. Compared with patent protection, this feature of instantaneous protection favours both copyright and trade secret protection. Compliance cost comparisons are even more favourable to copyright. Roberts has estimated that application fees and costs for patent protection amount to more than $1500 on average and estimates that these costs are negligible for copyright protection. 26

3. **Enforcement Costs of Copyright**

The history of copyright has established reasonably clear rules about what constitutes infringement. Certainly translating a program from one computer language to another would constitute infringement. Also, transcribing the program from one medium to another would constitute infringement. In this latter case, the analogy to music, perforated piano rolls, and sound recordings is clear. Just as transcription from sheet music to piano roles would constitute infringement, 27 so should transcription of software from printed form to magnetic tape or read-only-memory on a silicon chip.
One of the techniques currently in use to facilitate enforcement of trade secrecy in software and of copyright in directories is the insertion of meaningless statements or data in the original. In a computer program, statements can be inserted which the computer will never get to and never execute. In directories, fictitious names and addresses can be inserted. In both cases, infringing copies which have slavishly copied the original will contain the meaningless statements or fictitious entries. The infringer cannot then claim that his work was the result of independent effort on his part, for it is unlikely that in his independent effort he would have included precisely the same irrelevant material. That such techniques are available for both copyright and trade secrecy enforcement suggests two important and related points. The first is that both forms of protection pose enforcement difficulties. The second is that independent effort can form part of a defence in each case.

Neither copyright nor trade secrecy grants a preemptive right of any sort. If someone can demonstrate that they wrote a computer program independently but which happens to have numerous similarities with a program written by another person, she will not be guilty of either infringing a copyright or violating a trade secret. Major differences between the two types of protection in terms of enforcement are that under trade secrecy, access to the secret must be established and that under copyright, damages are recoverable from infringing third parties. In practice, these differences are not great, for, in the former, access is usually an important element of a copyright case; and, in the latter, the law of torts can be applied to third parties which encourage the violation of trade secrecy.
The discussion of enforcement thus far suggests that copyright offers little advantage relative to trade secrecy. The differences are relatively minor, suggesting, as has already been indicated, that the expected net social benefits of copyright protection vis-à-vis trade secrecy would be small. These minor benefits, however, would be the result of the creation of additional and/or more valuable and/or more widely disseminated software. Regardless of the source of the social benefits, they would likely occur in conjunction with additional enforcement costs, if for no other reason than that additional programs would be available. The paucity of cases involving trade secrecy or copyright and computer software, especially in Canada, makes it difficult to estimate these additional enforcement costs. Their non-existence does indicate, however, that the small increase in the number of programs would be likely to result in very few additional suits and impose small additional social costs in the form of enforcement costs. And to the extent that copyright enforcement might be cheaper than trade secrecy enforcement, the additional costs resulting from enforcement of more cases would be ameliorated by lower enforcement costs in each case.

4. Other Problems with Copyright

a. Term of Protection. It has been argued in this paper that software and especially firmware are analogous to music and sound recordings. Following that analogy, it seems reasonable that if copyright protection were extended to include computer software, the term of protection would be fifty years from the date of creation rather than the longer term of life of the author plus fifty years. Fifty years is a long time to grant intellectual
property protection for computer software, which by even the longest estimates has a durability of no more than twenty years. If, however, no program has a durability of greater than twenty years, then a fifty-year term of protection is unnecessary but it is also irrelevant, having zero social costs and zero social benefits for the period beyond twenty years. Comparing the term of protection with the durability of software is, at any rate, an incorrect approach to setting the term of protection. The optimal term depends not on what the current state of the art is but on what it will be or could be in the future and on what incentives society wishes to create. We could create a term of protection of five years if we wished, and that would create an incentive to produce short-lived software. By setting a longer term of protection, we would be offering a greater incentive to work on software with greater durability. The shorter terms would also encourage firms to eschew copyright protection in favour of trade secrecy in the hopes of maintaining monopoly control of their property beyond the term of copyright protection, whereas a longer term of protection would encourage less use of trade secrecy and greater dissemination of copyrightable works. Traded off against these benefits of longer terms of protection are the social costs of granting extended monopoly protection unnecessarily to works which would have been created even if the term of protection had been shorter. Balancing these benefits and costs is extremely difficult. Articles on the optimal life of patents abound, and their discussions can easily be applied to copyright protection as well. The major difference is that the monopoly granted by copyright protection is extremely limited, and, consequently, the social costs of a term of protection which might be longer than necessary to elicit creative works would
be extremely small. At a minimum, one can defensibly argue that the term of protection should be at least twenty-five years for two reasons: (1) to encourage longer-range development and wider-range dissemination; and (2) to satisfy Canada's obligations under the Universal Copyright Convention.

b. International Obligations. Canada is a signatory to two major international copyright conventions: the Berne Convention and the Universal Copyright Convention. It is reasonably clear that the international copyright conventions do not require the protection of computer software. It is also reasonably clear, however, that any member country which does extend copyright protection to computer software must do so for the nationals of all member countries.

To this point in the discussion, Canadian software producers have been treated as if they operate in isolation from the rest of the world. Such is not the case. Some Canadian firms produce and market software for customers in many countries, and some Canadian customers purchase software which has been developed and marketed primarily in other countries. Once these international trade flows in computer software are taken into account, the estimates generated so far will require qualification.

If there were free trade in a world market for software, Canada would certainly not benefit from extending copyright protection to computer software, and it might well be considerably worse off than it would be if it could expropriate software from other countries without fear of retaliation in other areas, and crucial "if" discussed by Whalley. But to demonstrate the conditions under which copyright of software in Canada would not generate any benefits for Canadians, a number of highly restrictive conditions would be required. When
these conditions are relaxed, it becomes clear that copyright could generate some benefits. What cannot be established, however, due to the lack of data, is the size of the benefits (or more precisely, how much the earlier estimate of these benefits should be reduced). In the first place, there are some barriers to international trade in computer software. There are tariffs and/or other import restrictions in many countries. Furthermore, communication and transportation costs contribute to limiting the market from a world scope to continental, if not national, boundaries. Finally, due to different legal, social, and cultural environments, the demands for different types of computer software will vary across national boundaries, contributing even further to the creation of national as opposed to broader markets. The effect of these forces is to make the Canadian market more relevant for Canadian software producers than would otherwise be the case; and to the extent that this occurs, benefits can be created for Canadians by enacting copyright protection for computer software.

In the second place, it may be stretching reality to say that the Canadian market has a negligible impact on decisions made by software producers throughout the world. At the margin, being able to protect the intellectual content of computer software in Canada will affect some decisions about development, marketing, costs, and prices, and it will consequently generate some benefits for Canadians.

In general, then, we are forced by this analysis of international considerations to acknowledge that a minimum of estimate of the area of the trapezoid in Figure 4 is probably less than $40,000, but how much less (and even whether it exists at all) we are unable to state.
Let us consider, as an example, the possibility that only ten percent of the benefits measured earlier would actually be created by Canadian copyright because Canada is such a small portion of the world market. In this case, the social benefits from copyright would be at a minimum $4,000 per year, but the administrative costs could still be as high as $20,000 per year if there are indivisibilities in the administration of sections of the act. In this case Canada would lose $16,000 per year. Alternatively, if one uses the upper bound estimate of the benefits, ten percent of these is still $200,000 per year, generating substantial net gains for Canada.

These numbers pose an interesting problem: It is difficult to reject two competing hypotheses with any confidence. If one wishes to establish that the net benefits would be positive, one is bound to examine the lower bound of the estimate of the benefits, and on balance this lower bound could be as low as -$20,000 or even considerably lower, indicating the difficulty of establishing the existence of these benefits. Yet if one wishes to establish that the net benefits would be less than zero, one should then examine upper bounds of the estimates; and plausible estimates of the upper bound seem to be certainly greater than $100,000 and perhaps as much as $2 million per year, not even considering the potential gain in the area of firmware and other software in the growing microcomputer market. In neither case can the hypothesis be rejected. In the former it is difficult to reject the hypothesis that the net gains would be negative. In the latter it is difficult to reject the hypothesis that the net gains would be positive.
The policy question, then, is this: should the Federal Government spend what would probably amount to less than $10,000 per year to obtain the potential gains from making computer programs copyrightable? My answer to this question is yes. While it is possible that the gains would be less than $10,000 per year, and Canada would suffer a loss, I suspect that they will, especially in the future, be considerably greater than $10,000. In my view, the investment of $10,000 per year in administrative costs would be a good risk for Canada because the potential gain from this investment is great relative to the potential loss.
Footnotes

1 Harold Demsetz, "Information and Efficiency," JLE, April, 1969.


4 Ibid., p. 20.


6 Several qualifications to this rather strong statement are important. For one, rent-seeking behavior by potential competitors may erode the private benefits and yet create a social deadweight loss. (2) It is conceivable that externalities could be generated which would reduce private benefits and leave a substantial social loss. (3) The analysis uses consumers surplus, a concept which, though often used in analysis of this type, is not always applicable. And (4) the analysis avoids the general problem of second best, i.e., it examines only one industry in the economy without considering feedback effects from one industry to another.

7 Miller, n. 194, supra, p. 59. The survey reported by Miller, we hasten to caution, has a wide margin for error. It was quite small in scope and covered primarily larger firms which might be expected to have the most to gain from intellectual property protection of software.

8 Miller, n. 195, p. 25.
Ibid., p. 27 (emphasis in the original). As Link has pointed out to me privately, interpretation of these data must be cautious. Perhaps many people were all aware of one incident.

Ibid., p. 22.

Ibid., p. 28.

12 This preliminary discussion has not mentioned computer firmware, a segment of the industry discussed in detail in my earlier study, Copyright and the Computer, Copyright Revision Studies, Consumer and Corporate Affairs, Ottawa, 1982. At this point I only wish to note that the problems of traditional trade secret protection for products as general and marketed as widely as firmware are possibly even greater than for proprietary software packages.

13 Of course, copyright will also create a limited monopoly right which could tend to inhibit a downward movement of prices. As throughout this study, however, this monopoly right would be extremely limited, protecting only a unique expression of an algorithm; others would be free to use the same algorithms in their own independently created expressions.

14 Diagrammatically, this effect is shown in Figure N.1 as the entire shaded area due to both the cost reduction and the increase in demand. Practically, however, the effect may be unimportant if the movement from \( Q^* \) to \( Q^{**} \) is assumed to be due solely to cost reduction. In this case, for a given price elasticity of demand, the supply curve would be assumed to shift downward, not to \( S' \), but to \( S'' \) and a larger trapezoid would be estimated. Although there is no reason to expect this larger trapezoid to be equal in area to the shaded area shown in the figure, it will generally be a reasonably close approximation. To the extent that these areas represent a transfer of economic activity from black markets to white (?) markets, they overstate the net benefits.
Figure N.1
15 Recall that registration is required for copyright protection in the U.S.


18 Data provided by Andre Gariepie, Commissioner of Patents, August 1980.


20 Gemignani, n. 10 supra, p. 301.

21 These numbers were mentioned by Miller, n. 194 supra.

22 See Frank, "The New Software Economics," Computerworld, Jan. 22, 1979. These calculations are carried out as if all software is the large-system, proprietary type. The impact of copyrights for microcomputer software would undoubtedly increase these figures. See Grant Buckley, "Legal Protection for Software Still Foggy," Computing Canada, Nov. 25, 1982; and Dave Ferris, "Piracy is the Software Industry's Achilles' Heel," Software News, August 2, 1982.

23 But again, remember that the price data refer to proprietary packages only. The prices of microcomputer software are much smaller.

24 Data provided by the Office of the Registrar of Copyright.
25 By 1982, with copyright protection clearly available in the U.S., registrations were running at a weekly rate of 70! Examiners there are alleged to be able to handle 40 applications per day, so that surely the cost estimates presented here are reasonable. See John Dodge, "Argue Over Software Copyright," Software News, April 5, 1982.

26 See n. 19 supra.

27 The White Music decision in the U.S. has effectively been reversed and changed through new legislation and judicial reinterpretation.

28 See, for example, the references cited in J. Palmer, "Copyright and Computer Data Bases," International Rev. of Indust. and Intell. Prop. (1982).

29 See, for example, the extended discussion in F. M. Scherer, Industrial Market Structure and Economic Performance, Rand McNally (1980).

30 Keyes and Brunet, Copyright in Canada, 1976.


PERSONALITY AND PERSONALITY: THE RIGHT OF PUBLICITY

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Practical Aspects of the Right of Publicity</td>
<td>2</td>
</tr>
<tr>
<td>American Cases</td>
<td>4</td>
</tr>
<tr>
<td>Canadian Cases</td>
<td>8</td>
</tr>
<tr>
<td>English Cases</td>
<td>10</td>
</tr>
<tr>
<td>The Property Issue</td>
<td>12</td>
</tr>
<tr>
<td>Economic Aspects</td>
<td>29</td>
</tr>
<tr>
<td>Conclusion</td>
<td>38</td>
</tr>
</tbody>
</table>

Howard P. Knopf  
Barrister & Solicitor,  
Suite 1400,  
180 Bloor St. W.,  
Toronto, Ontario,  
M5S 2Y6

© Howard P. Knopf, 1983
INTRODUCTION

Personalty is personal property, as contrasted with realty, or real property. It includes, of course, intellectual property. One of the most fascinating areas of development in the law of intellectual property deals with the concept of claims in a name, likeness or personality. Although American law is rapidly heading in the direction of enshrining these claims as "rights" into the annals of intellectual property law, I propose to demonstrate in this paper that such a development would lead to significant legal and economic policy problems. Accordingly, my thesis can be stated rather simply: claims in a name, likeness or personality give rise to a cause of action in tort during the holder's life only, but are not property rights in any known or reasonably conceivable sense of personalty.

"What's in a name?" Who shall inherit fame? These are really the basic questions in the legal issue of personality, or as it is most often called, the right of publicity. Legally, the answer has been and still is very unclear. But the marketplace has recognized rights in these commodities as extremely valuable. Moreover, the death of such marketable personalities as Groucho Marx and Elvis Presley have generated some rather important litigation regarding the "descendibility" of such rights.

From an economic viewpoint, these claims are valuable because a well known personality can be used commercially for such obvious purposes as the endorsement of products, or even as products in themselves (i.e. naming of restaurants such as "Winston's", "Gable's", or "Marilyn's"). Even more subtle uses include the use of a famous actor who has become closely associated with a fictional personality to convey or suggest a type of endorsement. For example, Robert Young endorses Sanka Coffee but the public may see this as a medical endorsement because of Young's long-standing role as Doctor Marcus Welby, a fictional character. Does Robert Young, or the copyright holder of the Marcus Welby story, or indeed anyone own the right involved here? Or what of the case of an unmistakable (yet unidentified) voice of a long-since
dead person singing a song now very commonly associated with a famous product? Does this situation involve any property right other than the obvious copyrights in the music and phonograph recording?

As we shall see, the "right of publicity" as a right and the "(mis)appropriation of personality" as a tort are the most common articulations of the concept we are dealing with. They are also, interestingly enough, apposite or converse expressions of the same phenomenon. The former implies a "right". The latter implies a cause of action. The former phrase is more commonly found in American jurisprudence, and the latter is the rubric most often used in Canadian and English discussions of the concept. In general, throughout this paper, these phrases will be used according to the jurisdictional context of the law or concept under discussion. In a more generalized context, I will use my own coined phrase "claim on personality" because it is less value-laden and less conclusory than the former phrases.

PRACTICAL ASPECTS OF THE RIGHT OF PUBLICITY

Whatever the legal basis may be for the right of publicity in Canada and the United States, the following are some practical aspects which are associated with the so-called right in routine commercial activity on the continent.

Reputable advertisers of products and services will not knowingly use any living person's name, likeness or personality in connection with any type of advertising without that person's consent.

The cost of this consent, in monetary terms, can be staggeringly high. It may be very unclear in the parties' minds why the figures are being paid, but nonetheless they are asked and given. A highly celebrated actor or athlete involved in a commercial campaign for a major product can easily earn a six-figure fee for a combination of his acting services, time, endorsement and use of his name, likeness and personality in connection with the campaign.

Laws relating to misleading advertising have made reputable
advertisers very wary of fabricating endorsements, and elaborate acknowledgments are usually executed by celebrities indicating that to a certain minimum extent they actually do use the product involved, and believe in it, quite apart from the substantial fees they are getting from so stating in public.

When it comes to the commercial use of a deceased person's name, likeness or personality, the practice seems to be less clear. For example, at the time of writing this article, IBM is running commercials for their Personal Computer on television and print advertisements featuring a Charlie Chaplin look-alike acting in a manner very similar to that of the celebrated actor's role in "Modern Times". Counsel for IBM has advised that although IBM was not convinced that consent of the Chaplin Estate was necessary, IBM made a last-minute decision to "play it safe" and obtained and paid for consent for this usage.

In a similar vein, but with an important difference, is a recent commercial campaign by Sony featuring an Albert Einstein look-alike, sitting in outer space and using a Betamax V.C.R. The important distinction from the Chaplin situation is that Einstein was not a professional "entertainer" and never marketed his "personality" as such. He was merely one of the century's greatest celebrities and certainly one of history's most distinctive persons in terms of his "likeness". Chaplin's widely remembered public image was a created character, albeit created by himself. Sony did not license this usage from Einstein's estate because they did not feel any legal obligation to do so.

Other areas involving problems in the law of personality include franchises, trademarks, naming of corporations, fictionalizations (i.e. an unauthorized biography containing untrue yet non-defamatory matters\(^1\)) and simply unauthorized biographies. All of these involve use of a person's name, likeness or personality for direct or indirect commercial purposes.

---

1 See Spahn v. Julius Mensner Inc. 21 N.Y. 2nd 124 (1967)
AMERICAN CASES

There are probably by now several hundred American court decisions dealing with the "right of publicity" as it is usually called in the U.S.A. To find a common thread of reasoning throughout is unfortunately impossible. The actions lie in quasi-contract, passing off, unfair competition, defamation, invasion of privacy, unjust enrichment, right of privacy and even copyright and trademark law.

Leaving aside for a moment the history and theory of the evolution of right in the United States, it appears impossible at this moment in time to even state the nature of the right. Moreover, although the nature of the right or claim itself is still far from clear, the current focus of the high profile litigation seems to be concerned with the "descendibility" of the right.

The flood of confusing American jurisprudence began with a trickle in 1890 in an article entitled "The Right of Privacy" by Warren and Brandeis in Harvard Law Review.¹ The next major milestone was a decision of Judge Jerome Frank in Haelan Laboratories v. Tops Chewing Gum.² This case held that the right of publicity existed in addition to and completely independent of the established right of privacy.

As to the right of privacy, Prosser has apparently definitively enunciated four distinct torts subsumed under the common law right of privacy:

1. Intrusion upon the plaintiff's physical solitude or seclusion;
2. The public exposure of embarrassing private facts about the plaintiff;
3. Publicity which would place the plaintiff in a false light in the public eye;
4. Appropriation, for the defendant's benefit or advantage, of the plaintiff's name or likeness.³

¹ 4 Harvard Law Review 193 (1890)
² Haelan Laboratories v. Tops Chewing Gum 201 F 2nd 866 (2nd circuit) (1953)
³ C.W. Prosser, Restatement (2nd) of Torts, Section 117
The next milestone after Haelan was the one and only thorough analysis of the U.S. Supreme Court in this area, namely Zacchini v. Scripps-Howard Broadcasting. The only other dealings of the U.S. Supreme Court in this area have been denials of certiorari. The Zacchini decision is much better known as the "Human Canonball Case". Mr. Zacchini had a daredevil type act which apparently consisted of being shot out of a canon. A broadcaster filmed the entire act and presented it as a news item. Mr. Zacchini sued on the basis that his right of commercial exploitation had been appropriated. The broadcaster responded that it was a newsworthy event and in presenting it, even in full, the broadcast was privileged under the United States first amendment law. Zacchini was successful in the Supreme Court, and the floodgates have never closed since. I tend to agree with Professor Sims that the Zacchini decision is probably limited or should certainly be limited, to certain circumstances in which a performer's entire act is appropriated and only compensation is sought, as opposed to injunctive relief.

The main thrust of the recent American litigation having most commercial importance concerns the descendibility of the late Elvis Presley's right of publicity. Up until quite recently, it seemed that the law of the State of New York held that the right was descendible and the U.S. Supreme Court refused to review the New York Appellate Level Court's view. See Factors etc. v. Pro Arts Inc. On the other hand, the U.S. Supreme Court also refused to review the holding by the Tennessee Appellate Level Court that the right was not descendible. See Memphis Development Foundation v. Factors. The seeming contradiction is based upon the apparent position that the right of publicity is a concern of each state and not one of federal jurisdiction. However, due to jurisdictional and conflict of law issues, the New York courts were obliged to apply Tennessee law in the Presley litigation.

1 433 U.S. 526 (1977)
3 579 F 2nd 215 Cert denied, 440 U.S. 908 (1979)
5 652 F 2nd 278 (2nd Circuit 1981) at 282
This, however is apparently subject to some serious questions in itself, since some writers thank that it should be subject to federal pre-emption.\(^1\)

To make matters even more confusing, the issue is now again very much at large in the State of New York \(^2\) because the State Court of Tennessee has recently concluded that the right of publicity is descendible in Tennessee.\(^3\)

Recent developments concerning Groucho Marx are also very important. A successful Broadway musical entitled "A Day In Hollywood/A Night in the Ukraine" was described by its authors as a "satyric comment" on Hollywood in the 1930's. The latter part of the play is a shameless exploitation of vintage Marx Brothers style, complete with unmistakeable imitations of costumes, faces, mannerisms, likenesses, voice personalities, and virtually everything except the names of Groucho, Chico and Harpo. None of the actual material was in any way alleged to be an infringement of any copyrighted material in any "genuine" Marx Brothers movies or scripts or other protected material. At the trial level, the plaintiff Groucho Marx Productions was successful in restraining the production of this play on Broadway.\(^4\) However, at the appellate level the courts of New York have just recently reversed this decision relying heavily upon their interpretation of California law arising out of the Rudolph Valentino and Bela Lugosi litigation. William, J. speaking for the Second Surrogate Court of Appeals, states:

"We conclude that Lugosi is subject to two interpretations. It may mean that California does not recognize any descendibility right of publicity and that the heirs of a celebrity must rely on trademark law to protect the goodwill that the celebrity brought to a product during his lifetime. Alternatively, Lugosi might mean that, wholly apart from trademark law, California recognizes a descendible right of publicity that enables the heirs to prevent the use of a celebrity's name and likeness on any product or service the celebrity promoted by exploiting the right of publicity during his lifetime... even if there is a limited descendible right, applicable to a product or service promoted by the celebrity, the defendants are not using the name or likenesses of the Marx Brothers in connection with any product or service that the comedians promoted during their lives." \(^5\)

---

2 Factors etc. v. Pro Arts Inc. 541 F. Supp. 231 (May 11, 1982)
3 Commerce Union Bank v. Coors, 7 MED. L. RPTR. 2204
4 Groucho Marx Productions v. Day and Night Co. 51 L.W. 2199
5 51 LW 2199 (9/10/82)
If one should seek clarification of the status of the law in New York by going back two years to the Lugosi decision,¹ one will not find too much comfort. This decision really consists of four opinions which analyze the right of publicity as a property right, a privacy right, an employer-owned product of employment, and a proprietary right which should be treated like a copyright.

The Lugosi decision has provoked some rather fascinating case comments. Kevin Marx, writing in the California Law Review² concludes "that personal and societal interests in the right of publicity cases are best advanced under a common law scheme that parallels copyright law."³

David R. Ginsberg analyses the Lugosi decision⁴ from a personal property and testamentary standpoint in such a way that I suspect that creative Beverly Hills attorneys are now contemplating provisions for the devolution of certain facial expressions of their prominent clients to the appropriate beneficiaries.

The foregoing is only a drastically simplified sketch of both the base and the tip of the iceberg which comprises the American jurisprudence pertaining to the right of publicity. One can find considerable jurisprudence dealing with publicity rights associated with totally fictional characters as portrayed by live actors, fictional cartoon characters, real voice imitations of other real voices, unauthorized but not quite libellous biographies, totally fictionalized accounts of real people and probably even animals. At one point, the heirs of Robert Schumann, the well-known and thoroughly insane 19th Century romantic composer, tried to assert an inherited right of publicity in the courts of New York approximately one-hundred years after their famous ancestor had died.

Having the naive security of a Canadian's viewpoint of American law on the subject (especially as constrained by the limitations of this paper) I can only conclude that a narrow view of Prosser's fourth tort classification, namely

1 Lugosi v. Universal Pictures 603 P 2nd 425 (December 3, 1979)
2 1970 California Law Review 786
3 Ibid 787
"appropriation for the defendant's benefit or advantage of the plaintiff's name or likeness" is the only generally accepted application of the right of publicity in the United States at this time, and then only with any certainty during the plaintiff's lifetime.

Several American states do have legislation dealing with privacy although this legislation does not figure prominently in any of the major jurisprudence.

**CANADIAN CASES**

Unfortunately, space limitations in this paper preclude me from dealing with Canadian case law other than in passing. The only cases of any significance with regard to the claim on personality are *Krouse v. Chrysler* and *Athans v. Canadian Adventure Camps Ltd.* For the sake of completeness the reader is referred to *Racine v. CJRC Radio Capitale Ltee.* and *Heath v. Weist-Barron School of Television Canada Ltd.* The latter decision was merely an interlocutory proceeding allowing this particular case to go to trial, but I am advised by counsel for the defendant that at the time of this paper it is proceeding to trial on an undefended basis because the defendant has gone out of business.

For an excellent review of the Canadian case law on this subject, the reader is referred to Professor Irvine's article. Furthermore, Professor Vaver has done a very thorough analysis of privacy legislation in British Columbia, Manitoba and Saskatchewan relevant to this area.

Although there is an extremely limited body of law in Canada dealing with the claim on personality, it is still difficult to extrapolate the present status of Canadian law on this subject. Nonetheless, the following principles seem to have emerged to date:

1. There is, in principle, a tort of wrongful appropriation of

---

1. Prosser, Op. Cit. at 814
3. (1977) 17 O.R. 2nd 425 (High Court of Justice of Ontario)
4. (1977) 35 C.P.R. 2nd 236
5. (1981) 34 O.R. 2nd 126 (High Court of Justice of Ontario)
another's personality for commercial purposes.

2. In order to succeed in any action, it will be necessary to prove actual damages. This could be done in a clear cut case of an unauthorized endorsement by a celebrity who has been in the habit of marketing his personality.

3. In other cases, the outcome under Canadian law at present is most unclear.

4. Canadian law has shown no interest to date, other than very distant obiter dicta, in any theory of property, copyright or trademark origins of this claim, and eschews the term "right of publicity".

5. There is no meaningful discussion whatsoever to date in Canadian law of such issues as descendibility of the claim on personality, although the British Columbian and Saskatchewan legislation provides that it does die with the plaintiff. Although under Manitoba law, an action for commercial appropriation may be continued by the deceased plaintiff's estate, this does not necessarily indicate that a plaintiff's estate in that jurisdiction would be able to sue for an appropriation of the plaintiff's name occurring after the death of the plaintiff.

The Province of Newfoundland has also passed an act called The Privacy Act. It should be emphasized that the small body of Canadian case law referred to above would not necessarily apply in the four provinces in which "privacy" legislation has been passed.

---

1 See Privacy Act, RSBC, 1979, Chapter 336, Section 5 and The Privacy Act, R.S.S. 1978, Chapter P-24, Section 10


3 Statutes of the Province of Newfoundland, 1981, c. 6
ENGLISH CASES

The common law of England remains to this day the most important influence on Canadian common law. Therefore, it ought to be of great significance to Canadian lawyers concerned with the matter of the claim on personality that English courts have stopped somewhat short of even recognizing that it exists. Recently, doubt was expressed in an interlocutory proceedings as to whether there was a real prospect of the pop group ABBA succeeding at trial in preventing the use of their name and likeness on T-shirts and other merchandising.\(^1\) See also Harrison v. Polydor\(^2\) where it was held that there was no real prospect of preventing the use of photographs of the Beatles on record sleeves containing recordings of interviews with them. In fact, a respected recent English text writer cites the Krouse\(^3\) and Athans\(^4\) cases as the leading commonwealth cases on the subject and states:

"No English court has been prepared to go even so far, let alone to address the question whether an unknown person can object to the use of his "personality" (name, photograph, etc.) in commercial or other propaganda without his permission... the common law has remained unwilling to prevent all unwanted personal publicity under the rubric of "a right of privacy" or a "personality right" as we have noted, there must be some more specific form of injury: for instance, through breach of confidence or a defamatory statement."\(^5\)

The same writer notes that there is a good case for holding that a person's name, appearance or voice in commercial publicity without his permission should be actionable.\(^6\) However, he states that the current legal status of the issue is unsettled in England and the fact that permission is sought and payment is made is essentially a question of advertising ethics and voluntary practice.\(^7\)

I do not propose in this paper to examine English case law as it applies to the Canadian or North American context in more than a passing manner. There is an excellent analysis of it in this context in Professor Irvine's article.\(^8\)

---

1 See Lyngstad v. Anabas 11977I F.S.R. 62
2 11978I F.S.R. 1
3 1 O.R. 2nd 225
4 (1977) 17 O.R. 2nd 425
6 Ibid, Page 492
7 Ibid, Page 492, 494
However, the following broad generalizations should be stated about English law, for present purposes.

Virtually all possible theories pertaining to the claim on personality have been canvassed by the learned English judges since as early as 1816.¹ By 1870 it was appearing to be the case that the English jurists were tending to treat a person's name in equity as a species of property. This was later discredited to a large extent and the mainstream of English cases since then deal with principles of defamation or passing off, which are very ancient, honoured and reasonably well-understood and immutable concepts in English law.

¹ See Lord Byron v. Johnson (1816), 35 E.R. 851
THE PROPERTY ISSUE

My very cursory examination of relevant case law above was intended to show that the claim on personality as it currently stands cannot be neatly categorized as a clear-cut property right on the one hand or a right, the violation of which gives rise to a cause of action purely in tort, on the other hand. Although in many particular instances, this particular distinction and dilemma may not be relevant in achieving justice, I believe that confusion in this area will only be removed and sound jurisprudence achieved once a decision is made as to which theory is definitive. At least in Canadian and English jurisprudence, the decision would still seem to be open for full consideration.

Notwithstanding that I am addressing an audience comprised mainly of intellectual property lawyers and economists, my argument is that the tort theory should be adopted and the property theory should be discarded. In this section of my paper, I propose to justify this thesis from a legal and practical viewpoint. In the following section, I will put the issue to an economic analysis test.

For purposes of convenience, the terms "name", "likeness" and "personality" will hereinafter be referred to as "personality" unless otherwise specified. First, I wish to attempt to define personality. I would suggest the following:

"A person's personality is the minimum set of characteristics which identify him in the perception of the public as a particular and unique person apart from all others".

The key concept in this definition is that of "minimum set". For example, many of people have moustaches, wear glasses and smoke cigars. If the public, when confronted with these three particular aspects, generally would call to mind one Mr. Groucho Marx, Esquire, then one must give Groucho credit for having established a kind of informal "trademark" out of these simple commonplace elements in combination. However, the appearance of these three aspects without further refinement is by no means unique and many more precisely defined aspects are required to delineate the "distinctive"
Groucho personality in the technical sense of my definition, so as to exclude other persons who happen to share these three aspects.

In some particular personalities, it may be one element as simple as the particular "frog" in Louis Armstrong's voice that calls his identity to mind in the eyes of the public. Hopefully, my definition is reasonably precise and does not permit absurd conclusions to follow, because I would not care to advocate a definition which would allow person A to sue person B simply because person B may happen to share certain aspects of person A's personality in a context where this is trivial or irrelevant. If this were so, our learned brothers of the bar in California would soon be licensing hairstylists in Toronto for using the Farah Fawcett hair-do.

One could argue that artificial personalities, in a creative and non-pejorative sense, should be rewarded in the same way that originality is rewarded pursuant to copyright and patent law. On the other hand, one suspects that certain particularly unique personalities are in fact quite spontaneous and it would be hard to argue that these personalities should be less deserving of protection. Woody Allen and Albert Einstein come to mind as such personalities.

My definition could include totally fabricated and fictional dramatic personalities which are held out as owned, held or asserted by particular real people. For example, Robert Young is probably more perceived as the marvellous Doctor Marcus Welby then he is as Robert Young, Esquire. The makers of Sanka Coffee undoubtedly realize that the personality of the real (or apparently real) Robert Young and the totally fictional Marcus Welby are virtually identical in the eye of the public and that the public thinks that this particular actor is a real medical doctor. Presumably to counter this suggestion, Sanka later included in these commercials a disclaimer from Mr. Young that he is not in fact a real doctor. Let us assume (although it is not absolutely clear) that the owners of the copyright in the scripts to the Marcus Welby series could sell outright the rights to the character, the name, the plots, scenarios and forms of expression associated with
the good doctor. They would not be able to exploit it further. But, does this mean that Robert Young must stop being Robert Young?

This, among other questions, become immediately problematical in any analysis of the question of personality and irrespective of any definition adopted of personality. Returning for the moment to the example of Doctor Welby, it is a very basic notion of personal property law that the so-called "owner" of property has the right and indeed the ability to dispose of or alienate the property in which he lays claim. He need not fully alienate it, but if he cannot do so it is difficult to understand what the property right involved really is. Robert Young cannot stop being Robert Young. On the other hand, more familiar types of intellectual property such as trademarks and copyright can, for all intents and purposes, be fully alienated. An interesting exception in the case of copyright occurs in the matter of the reversionary interest provided for in the law of Canada by Section 12 (5) of The Copyright Act, R.S.C., 1970 as amended.¹

It becomes necessary to examine some aspects of the meaning of personal property or personality. To begin with, the only practical definition of personality "is the slightly comic one that it is not really"². Moreover, no theory of property, whether real or personal, can avoid dealing with such issues as ownership and possession and that is also difficult. According to Sir John Salmond:

"So long as we remember that the ownership of a material thing is nothing more than a figurative substitute for the ownership of a particular kind of right in respect of that thing, the usage is one of great convenience; but so soon as we attempt to treat it as anything more than a figure of speech, it becomes a source of confusion of thought."³

¹ See Barry Torno, Ownership of Copyright in Canada, Copyright Revision Studies, 1981
₂ Page 107
⁴ Salmond, Jurisprudence (8th Edition) Page 279
On the question of ownership, Halsbury states that:

"Ownership consists of innumerable rights over property, for example the right of exclusive enjoyment, of destruction, alteration and alienation, and of maintaining and recovering possession of the property from all other persons. Those rights are conceived not as separately existing, but as merged in one general right of ownership."\(^1\)

The proponents of the property theory of personality often base their arguments upon an analogy with copyright.\(^2\) On the particularly troublesome notion of descendibility, they actually go so far as to state that "if the possibility of inherited publicity rights existing in perpetuity proves troublesome, an analogy to copyright law again suggests a solution."\(^3\) This, I respectfully submit, is ultimately bootstrapping of the most troublesome nature. It is true that copyrights survive generally for a period of fifty years following the death of the author. Patents have a considerably shorter lifetime and trademark rights can be lost quickly through improper licensing or abandonment. But who would have ever dreamed until very recently that a person's name, likeness or personality could be dealt with in a will or devolve as personal property or otherwise in any way? Ironically, American advocates (by copyright analogy) of the descendibility of the "right" seem to forget that until relatively recently in the U.S.A., a copyright could become public domain in 28 or 56 years, often before the death of the author!

Halsbury classifies copyright as personal property, along with "many kinds of property unknown to the common law" such as patents, shares, government annuities, etc.\(^4\) Interestingly, the English Copyright Act of 1956, S. 36(1) provides that "copyright shall be transmissible by assignment, by testamentary disposition, or by operation of law, as personal or moveable property." (Emphasis added).

\(^2\) See for example Kevin S. Marks, An Assessment of the Copyright Model in Right of Publicity Cases, Volume 70, California Law Review, 785 (1982)
\(^3\) Felcher and Ruben, The Descendability of the Right of Publicity; Is There A Commercial Life After Death?, 89 Yale Law Journal, 1124 at 1131 (1980)
Note that the English legislation does not deem a copyright to be personal property, but merely provides that it has most of the attributes commonly associated with personality. The Canadian Copyright Act is similar in its giving of attributes, (see Copyright Act, R.S.C. 1970, as amended, Section 12) but makes no reference to the words "personal property".

Crossley Vanes recognizes copyright as personal property because it is a chose in action.¹

Thus, it can be seen that although copyright is now treated for all intents and purposes in Anglo-Canadian law as personal property its pedigree and status as personality is still theoretically somewhat suspicious and uncertain.

Even the Supreme Court of Canada has lately pronounced on the anomalous status of copyright. Estey, J. in a recent decision observed in obiter dicta that:

"...Copyright is neither tort law nor property law in classification but is statutory law. It neither cuts across existing rights in property or conduct nor falls in between rights and obligations heretofore existing in common law. Copyright legislation simply creates rights and obligations upon the terms and in the circumstances set out in the statute."²

Returning to my concern for the bootstrapping element entailed in the treatment of personality as property, especially as a copyright, I would suggest that most basic notions of personal property law and tort law strongly mitigate against such a treatment. It is trite Anglo-Canadian law that there is no property in a person's

name per se. Although misappropriation of a person's name or likeness may indeed give rise under appropriate circumstances to a cause of action in personam as seen in several of the cases cited in this paper, this is not the same thing as copyright, of which Goodeve states "being an exclusive right in rem against all the world, it is to be considered as property in the author or his assigns". A fortiori, even less definable aspects of a personality (i.e. voice, hairstyle, mannerisms, etc) would appear to have even less of a claim to designation as property per se.

The perils of bootstrapping a phenomenon into the realm of intellectual property have been well summarized by Bruce C. MacDonald as follows:

"Few writings on the subject of intellectual property expose the circular and issue begging use constantly made of the word "property". "Property", of course, means little more than legal protection for a claim made by a person. It usually refers to the guarantee of entitlement to exclude. The reasons for finding such an entitlement necessary, in intellectual property law as in all other areas of law, an enquiry as to whether the conditions of protection are met. But whatever the precise definition of "property" the point here is that it is not reason to say that something deserves protection because it is "property"; "property" is a short-hand description for a conclusion of law. It is meaningless, for example, to claim protection on the ground that one has "natural property rights" in something. Land and moveable goods are commonly called "property" because they are typical subjects over which exclusive rights are recognized by law, but whenever the existence or extent of a right to exclude is challenged no assistance is gained by stating that one's interest is "property"." 3

Does it follow that even if the right of publicity is personal property, it is ergo descendible? It would seem to be a basic proposition of Anglo-Canadian law that "on the death of an owner his personal property devolves upon his personal representatives, subject to the administration of the estate, and the beneficial title is transferred in accordance with either his testamentary disposition or, in the case of an intestacy, the statutory provisions regulating the distribution

of the property of an intestate."¹ In other words, all personal property is descendible (except of course for successive interests subject to a trust not unlike a life estate)² and except of course for property which has perished before the death of its owner. It is also true (and trite) that all real property is descendible.³ Moreover all property is either realty or personality.⁴ Therefore, all property is descendible subject to the limited interest of its owner or perishability as stated above. This shall be our first premise for the following exercise in logic.

It now follows that, if all property is descendible, and personality is not descendible, then personality is not property. Nonetheless, on this premise alone, personality may be descendible without being property. However, it is empirically very difficult to think of any other remotely congruous example of such a status.

My second premise is that anything that is descendible is property.⁵ If, then, personality is not property, then personality is not descendible. But personality still could be property and not be descendible, if the second premise is in fact true. However this would clearly contradict our first premise based upon clear authority that all property is descendible, with the exceptions set forth above.

There appear to be three alternatives at this point. The first is that my second premise is incorrect, although empirically it would appear to be true. The second is that personality is a type of property akin to a life estate, successive or other limited interest, or perishable property. Personalities indeed appear to be perishable (e.g. those of Elizabeth Taylor and Orson Welles) or, indeed, the owner may perhaps hold only a life estate in them, but this is more of a theological than a legal issue. The third alternative is that a personality is simply neither property nor descendible and is thus like the air, the waters of the ocean, and a sunset.

Moreover, many of the particular manifestations of a personality are perfectly capable of protection in law by copyright or otherwise without the drastic legal leap of faith involved in making personality a property right. For example, a particular photograph of a celebrity will be protected by copyright, although not necessarily in the hands of that celebrity. If the photograph was taken in a public place, the photographer will own the copyright in the photograph, although improper use

² Crossley Vanes on Personal Property, 5th Edition, P. 31
³ Ibid P. 3 and P. 8
⁴ Ibid P. 3
⁵ This seems obvious, but oddly it is difficult to find an authority in so many words for this proposition.
of the photograph for purposes such as an unauthorized endorsement can be restrained under existing Canadian tort law. If the photograph was made for valuable consideration at the behest of the celebrity, the celebrity will own the copyright in it in the absence of any agreement to the contrary under Canadian copyright law.\footnote{Copyright Act, R.S.C. 1970, Section 12(2)} The photograph is, of course, a fixation of a person's likeness. Likewise a celebrity can provide contractually for ownership of a copyright in a particular fixation of his voice or an acting performance. If, through negotiating parameters, he does not own the copyright, he will in all likelihood stipulate for and receive royalties which may continue to accrue to his estate long after the particular celebrity has died. Thus, virtually all tangible creative products and manifestations of a personality can be protected under existing property law or by contract. Those that are largely unprotected under Canadian law such as performances by musicians, are left unprotected due to conscious policy decisions which may or may not be correct from an economic policy standpoint but are perfectly well understood.

Where the property-copyright analogy really breaks down, in my view, is in analysis of specific situations where one might seek to make it apply. The typical license agreement that purports to deal with the celebrity's name and likeness is clearly designed to accommodate existing case law which holds that the right must be exercised in one's lifetime in order to be descendible. A typical licensing agreement can be seen in \textit{Lindley on Entertainment Publishing in the Arts}, Volume 2, Page 1713 ff. These agreements normally state that they are not contracts for personal services and that the licensor-celebrity will receive additional compensation for any specific services rendered pursuant to the license. However, it can immediately be seen that the license will become an exercise in futility if the celebrity becomes disinclined to exploit or to allow the licensee to exploit his personality in general or in a specific instance. Although Professor Treece does cite one apparently very minor case in which
a defendant actually forced a plaintiff to appear in person to advertise the defendant's product, the basic legal notion that contracts for personal services are not specifically enforceable arose in the entertainment industry because actors and similar types of celebrities cannot be made to render a personal service if they are not so motivated. They certainly may be liable for damages if they capriciously refuse, but that is a different matter. How then can a contract licensing an alleged right of publicity be compared to an agreement licensing a copyright? Once a composer has written a song and sold or licensed it to a publisher, he cannot stop them from using it according to the terms of the written documents. On the other hand, an agreement by a composer to write music in the future is only as good for all intents and purposes as the composer's desire to fulfill it.

An interesting comparison can be made between the right of publicity or the claim on personality, and moral rights under copyright legislation. According to Professor Nimmer:

"It is sometimes said that moral rights as applied in Berne Convention countries are inalienable by their very nature since they constitute an element of the author's personality rather than a property right. However, not all countries which adhere to the doctrine regard it as inalienable and the Berne Convention does not require such inalienability." 2 (Emphasis Added)

This would appear to be an instance of an aspect of copyright that is not only owned by an author but cannot be alienated by him in certain jurisdictions. 3

Another interesting comparison with the notion of moral rights, especially since Nimmer would appear to view elements of a person's personality as inalienable, is that the duration of the moral right appears to be an uncertain question in itself. According to Barry Torno, under the present Canadian copyright legislation the duration is arguably perpetual. 4

---

1 Trees, 51 Texas Law Review, 637 at 639, Note 12
3 See Torno, Op. Cit., Page 72 FF
4 See Barry Torno, Torno on Copyright Protection in Canada: Present and Proposed Copyright Revision Studies, 1980, Page 36
Torno advocates that the term of protection for moral rights should be equal to the life of the author.\textsuperscript{1} It is suggested that it would be consistent with Mr. Torno's and Mr. Nimmer's viewpoints on the subject of moral rights that personality as a right should also terminate with the death of the "owner".

Thus, in my view, the treatment of the right of publicity or the claim on personality even as an incorporeal personal property right analogous to a copyright is both highly problematical in legal theory and totally impractical in terms of actual application. Moreover, and more to the point, the property theory of the right of publicity is leading, in my view, to conclusions that defy logic and common sense and is based upon false assumptions. Many courts and commentators are now suggesting, based on the property theory, that if a celebrity's right of publicity is dealt with during his lifetime in certain contractual ways, then that right should be descendible in the same way that any other piece of property can be devised, or by testamentary instrument.\textsuperscript{2} In an article apparently written at a time when the shifting winds of this issue were blowing in the direction of descendibility, one learned author suggested "consideration should hereafter be given to the disposition of this valuable and recognized right in estate planning for subject parties with substantial publicity value, lest it devolve to an unintended recipient either through a residuary clause or through intestacy".\textsuperscript{3}

The proposition that the right of publicity should be descendible in the same way as a shopping centre, car dealership or any other "developed" piece of property is usually defended by some variation of the "career incentive theory" which holds that:

1. The celebrity's belief that his publicity rights would be legally enforceable, hence valuable, after his death.

\textsuperscript{1} Ibid, p.37
\textsuperscript{3} David R. Ginsberg, Right of Publicity, 25 Copyright Law Symposium, ASCAP, 1980, P. 20
2. The celebrity's desire that the rights should survive in the hands of assignees, heirs, or other successors, and

3. that the celebrity's contemplation of survivability created added career incentive which

4. may have resulted in additional career achievements from which society benefited. ¹

In the celebrated Laurel and Hardy case² the widows of Laurel and Hardy sued to enforce a contract dealing with the late stars' right of publicity. They were successful, although it has been suggested that Laurel's case was really much better than Hardy's because he had made an assignment in perpetuity of his right at the age of 71, only four years before his death and more than ten years after his last professional engagement. On the other hand, Hardy had never made an inter vevos assignment in gross.³

It has also been suggested that descent of the right of publicity would contribute to give performers more incentive to make the investment required to produce a performance of interest to the public.⁴

Possibly the most fanciful lawsuit ever to reach the courts in this area was that of Schumann v. Loew's Inc.⁵ in which certain alleged great-grandchildren of the colourful and insane 19th century composer, Robert Schumann, unsuccessfully contended that a film dramatization of the composer's life violated the property right that had descended to them. The court apparently had no problem with the property cases upon which the plaintiffs relied, but dismissed the lawsuit on the basis that some one-hundred years had passed since the plaintiffs' dearly beloved

¹ Sims, Right of Publicity: Survivability Reconsidered, 49 Fordham Law Review, 453 at 472
² Price v. Hal Roach Studios Inc., 400 F Sup. 836
³ See Fletcher & Ruben, Op. Cit. at 1618-1619
⁴ Zaccchini v. Scripps Howard, 97 S. ct. 2849 (1977) at 2857. See also Lucosi v. Universal Pictures: Descent of the Right of Publicity, J.B. Eisenberg, 17 PEAL 311 at
⁵ 135 N.Y.S. 2nd 361 (Supreme Court, 1954)
ancestor had passed away. Clearly, this was a wise decision by the court because to have held otherwise might have invited lawsuits by heirs of Napoleon, Mozart and perhaps even Charlemagne. At the risk of stooping to the level of reductio ad absurdum, I would suggest that there is no rational length of time to uphold the existence of a right of publicity after a celebrity's death.

A dissenting opinion by the Chief Justice of California in *Lugosi v. Universal*\(^1\) concludes that "the right of publicity recognizes an interest in tangible property similar in many respects to creations protected by copyright...that body of law is instructive on the descendibility issue". The learned Chief Justice further concludes that the right of publicity should be recognized for a period similar to that of copyright, namely the creator's life plus fifty years.\(^2\)

How can I deprive a dead celebrity of the opportunity of doing anything? To say that I am depriving his heirs is a bootstrap type of argument based on the notion that his heirs are entitled to exploit his personality for gain. I have no doubt that well paid celebrities such as Johnny Carson (estimated income of $5 million per annum) or Bob Hope (estimated net worth of approximately $1 billion) will provide generously for their heirs in a manner which they see fit. They have realized and been compensated for their ephemeral (in the literal, not pejorative sense) talent in very substantial monetary terms during the course of their lives. I would suggest that unless they are composers or authors they leave nothing other than memories of their rather considerable talents, for which members of the public paid very well during their lifetimes and to which members of the public should be freely entitled after their deaths. Undoubtedly, contractual royalties will also pour in to their estates.

In fact, the career incentive theory, which bears a great resemblance to one of the basic theories advocated for copyright protection, simply does not bear scrutiny in many examples of commercial exploitation of personality.

---

1 603 P 2nd 425, at 444 to 446

2 Ibid at 446-447
On the contrary, many will argue that such prominent and well-paid endorsers as Bill Cosby (who seems to espouse the virtues of everything from Coca Cola to computers) are quickly cheapening and rapidly depreciating the aesthetic and even monetary value of their personalities by so doing. Highly celebrated, so-called "serious" actors such as Lawrence Olivier, Peter Ustinov and John Houseman are now earning immense fees during their lifetimes for endorsements, and this may well prejudice their marketability as prestigious thespians. However, it is their considered commercial career choice and must be respected, if for no other reason than the huge fees which they receive. Copyright law enables creators such as Igor Stravinsky to achieve a certain amount of well-deserved wealth and devise it to his heirs, but apart from publishing, mechanical income and performance royalties, Stravinsky did not have any other particularly tangible ways to make a living other than conducting. This is certainly not the case for actors and athletes who can earn tremendous fees from their profession and from endorsements during their lives, which is more than sufficient inducement for them to develop marketable personalities and career skills.

Now that I have attacked property law notions as a basis for the right of publicity, I wish to state that I do think there should be a claim on personality recognized in law during a person's life but for reasons of legal and economic policy, this doctrine should be supported by tort law, or the law of unjust enrichment or restitution. Although no-one has as yet perfectly defined a "tort" it is possible to state the purpose of tort law which purpose I suggest lends itself very well to the present problem:

"Arising out of the various and ever increasing clashes of persons living in a common society, carrying on business in competition with fellow members of that society, owning property which may in any of a thousand ways affect the personal property of others - in short doing all the things that constitute modern living - there must of necessity be losses, or injuries of many kinds sustained as a result of the activity of others. The purpose of the law of torts is to adjust these losses and to afford compensation for injuries sustained by one person as a result of the conduct of another." 1

Tort law gives rise to causes of action which may have rather substantial monetary value, but are not treated in the same way as personal property.

To give an example, if I am struck by a careless driver of an automobile, I may collect a very substantial amount of money from him or his insurance company. I may be able to receive credit from the bank on the basis of an expected settlement. I may, under certain circumstances, be able to assign my cause of action as a "chose of action". My family may be able to sue and collect damages as well. But I cannot imagine any possible scenario in which I could legally, much less commercially, sell my right to collect money in the event that I am hypothetically struck by a car in the future. If I die from the injury, my estate could carry on the action for losses suffered. My family could claim damages as a result of the income I can no longer provide, the loss of parental guidance and other direct and actual losses of services or commodities to which they would have been directly and personally entitled had the tort not happened.

Under Anglo-Canadian law, the cause of action for defamation dies with the plaintiff. The law of defamation and libel is completely based upon tort theory, even though important economic interests may be involved. Although a person's reputation may be very valuable, the law of defamation treats abuse of it clearly as a tort and not a property issue. Accordingly, a dead person's reputation has no legally recognized property value to his heirs, apart from business goodwill in some instances which is outside of the scope of defamation law.

It is highly interesting for the purposes of these arguments that English law in the early 1870's was tending to treat a person's name, in equity, as a species of property. However, the approach was overruled by 1878.

Misappropriation of personality or use of one's right of publicity is, in my view, a tort analogous to trespass, defamation, or passing off. By now, it is probably a recognized distinct tort under the common law of Canada and the United

1 See Trustee Act, R.S.O. 1980, S. 38
2 Ibid., S. 38
States, as seen above. In many particular instances, depending on facts, it might fall into the category of one of the more traditional sorts of torts, such as passing off. This could occur, for example, in the deliberate impersonation of a well-known celebrity for the purpose of a commercial endorsement of a product. Such a situation would appear to meet the recent and apparently definitive definition of passing off in the "Advocaat" case as stated by Lord Diplock:

(1) A misrepresentation (2) made by a trader in the course of trade (3) to prospective customers of his or ultimate consumers of goods and services supplied by him (4) which is calculated to injure the business or goodwill of another trader (in the sense that this is a reasonably foreseeable consequence), and (5) which causes actual damage to a business or goodwill of the trader by whom the action is brought of (in a quia timat action would probably do so).\(^1\)

In my particular example, the question of whether the goods or services are supplied by the endorser who is being impersonated may be problematical, but from a policy standpoint should not be so.

One admitted problem with limiting the approach to the claim of personality to tort law is that this doctrine, despite its remarkable flexibility over the ages, tends to be extremely conservative in the hands of English and Canadian judges. For example, the English courts do not really recognize the merchandising concept because they tend to view purveyors of T-shirts and dolls, etc., as being in a different type of business than those who make films and records,\(^2\) and despite the sometimes wide scope of the English common law in the area of passing off (i.e. the Advocaat case\(^3\)), English law is much more conservative in this area than American law (which tends to deal with such problems under the rubric of unfair competition or restitution) and even than Canadian law.

1 11979I F.S.R. 397 at 405
2 Wombles v. Wombles 11977I R.P.C. 99
3 11979I R.S.R. 397
One respected English writer has advocated that their law pertaining to merchandising should be changed, but only by Parliament because such a change would in effect grant a monopoly to the holders of the rights in the character being merchandised in the same manner as a copyright. The same writer also feels that "if such a change were introduced, it would presumably also include merchandising rights in the names and pictures of real people such as pop stars. As already suggested, their claim seems rather stronger."²

In one of the best American articles I have encountered in this area, written well before the post-Zacchini flurry of activity and the diversion into the descendibility issue, Professor Treece concentrates his attention mainly on the area of use of personality in advertising and endorsement situations. He concludes generally that "traditional tort remedies and injunctions undoubtedly apply to redress and prevent both economic and personal injuries caused by unauthorized advertising uses of names and likenesses."³ However, he then goes on to show that, even under American tort law, there can be problems in quantifying damages under tort law because damages can in many cases be purely nominal. See Flake v. Greensboro Newsco.⁴ He advocates an approach to damages predicated more on restitution or quasi-contract law.⁵ He suggests that, in the case of unauthorized use of a name or likeness for endorsement, a plaintiff who had been in the business of endorsing should be able to recover damages based on his previous fees and otherwise to establish damages on the basis of the going rate for celebrities of his nature.⁶

Although the notion of quasi-contract or restitution law is appealing as a refinement of the tort concept in this area in order to allow a plaintiff

---

1 Cornish, Intellectual Property, 1981 Page 500
2 Ibid, Page 501
3 51 Texas Law Review, 637 at 648
4 212 N.C. 780 (1938)
5 51 Texas Law Review, 637 at 649 F.F.
6 Ibid 651 - 652
recovery for the value of the benefit appropriated by the defendant, with all respect to Professor Treece, his method of establishing the "going rate" in actual cases may prove very troublesome. The actual figures in any given deal are usually a closely guarded secret. One must generally be aware of publicly "leaked" figures as they may tend to be inflated. Although most lawyers and agents in this area know full well that contractual fees for endorsement by celebrities are often in the six-figure and even higher area, they also know that these figures are negotiated very much on a case-by-case and person-by-person basis. Wayne Gretsky and Frank Sinatra and Brooke Shields are not exactly fungible property. The going rate concept would only likely work at the lower end of the scale where a person would be paid, if their services were used at all, on the basis of SAG rates in the U.S.A. or ACTRA rates in Canada.

My suggestion is that where a plaintiff of celebrity status is not able to necessarily put a more or less liquidated value on his loss of economic interest by misappropriation, he should nonetheless under the rubric of existing tort law, be able to recover on a punitive damage basis. Advertising agencies certainly are aware of the general law in this area and if their clients are not always so aware, it is probably time that the law forced them to be, at least in such clear cut areas as misappropriation for express endorsement purposes. In any case, restitution theory may not be necessary at this stage in the United States. Under American law, there is really no longer any doubt that the misappropriation of personality or abuse of the right of publicity is purely and simply an actionable tort and that leaps into the law of restitution and quasi-contract are no longer necessary. At minimum, there now is a tort of appropriation of personality or abuse of the right of publicity in both the United States and Canada under common law. There is no totally precise rule as to when it should apply and what the measure of damages should be. However, I have no doubt that that will evolve in time as long as the area can be focused into tort law and not diverted and diffused by concerns with copyright law, will drafting and descendibility.
ECONOMIC ASPECTS

What are the economic implications of the recognition of the claim on personality, whether founded in property law or tort? For the following analysis, I am indebted to a recent study prepared for the government of Canada by Messrs. Globerman & Rothman¹ (hereinafter referred to as Globerman). Mindful of the dangers of argument by analogy, I should point out that their study was concerned with performers' rights and in turn whether or not copyright legislation should be extended to provide a form of protection for re-use of recorded performances of performers such as actors, dancers and musicians. The analogy is, I suggest, closer than it appears. The law in Canada has never recognized, even in the most tenuous way, any copyright in the work of a performer. Once the performance has been done (whether recorded or not) the performer has been left solely to the enforcement of his contract for that performance. The only grey areas, with which we are not really concerned, are questions as to whether the content of certain performances, especially in such areas as dance and choreography, is so infused with the work of the performer, that a copyright might arguably exist in the performer's input.

Moreover, the re-use of the performance often does involve the use, abuse or appropriation (depending on one's viewpoint) of the performer's "personality" as I have defined it earlier.²

Globerman's analysis concerns itself generally with the economic implications to the performer after the performer has voluntarily placed himself in front of a camera or microphone. Except in certain cases applicable to the first scenario outlined above, the use of a personality is likely in most instances to be a re-use, adaptation or even, arguably, a gross distortion of a use of personality already authorized. An unauthorized "piracy" of a live rock concert, for example, is actionable in most cases on the basis of infringement of copyright in the music itself, or on the basis of breach of contract arising out of the terms of admission.

¹ Globerman & Rothman, An Economic Analysis of the Performer's Right, Copyright Revision Studies, Consumer & Corporate Affairs Canada, 1981 (herein referred to as "Globerman"
² Vide supra, page 12
or sponsorship. However, unless the performance itself is fit matter for copyright (i.e. in terms of originality and fixation) there may not be any other clear basis of actionability.

Notwithstanding the foregoing, I have one serious reservation about applying the Globerman analysis to the present subject matter, which is as follows. Globerman's analysis is focused on the economics of the average or median performer, who is not a great celebrity and who, at the present, does not make a great deal of money. In fact Globerman finds that the average income of full-time performers in Canada compares favourably with average annual income per person in Canada and that "the stereo-type of the starving artist is inappropriate when applied to professional performers in Canada".¹ This conclusion is, I suggest, based upon entirely obsolete 1971 statistics and as I hope to demonstrate below, their analysis cannot really apply to the "superstar". After all, the whole area of the right of publicity is mainly concerned with superstars whose names are familiar to most people.

The essential aspect of the Globerman analysis deals with the market structure and the existence of "rents" in the area of the performers' right. They define economic rent "as payments to a supplier above the amount required to encourage the existing amount of output in the long run."² The Globerman report analyses, as critical sectors in the marketplace being studied, the television broadcasting industry, the radio broadcasting industry, the film production, distribution and exhibition industries and other potential engages such as the record industry and performing arts companies. They conclude that a substantial case for the existence of rents can be made in the case of television broadcasters, a possible case for radio broadcasters and very little of a case for the others.³

On the basis of the rent analysis scheme, they analyse three main situations: (1) creating a permissive right where one already exists; (2) creating

¹ Globerman, Page 18-19
² Globerman, Page 56
³ Globerman, Page 65 F.F.
a specific right where a general one exists; and (3) creation of a re-use right where one does not exist. My analysis will be roughly parallel to this approach.

Before embarking upon this analysis, I respectfully submit that even within the scope of the Globerman analysis, two key areas are omitted in terms of a study of the existence or non-existence of rents in the system. These are the advertising industry (comprising advertisers and advertising agencies) and the consumer public who indirectly pay for all advertising and directly pay for consumption of performances at movie theatre box offices, through pay television subscriptions, video-cassette rental, etc.

Although I do not have any concrete evidence it would seem reasonably obvious that a very strong case can be made for the existence of substantial rents in the hands of advertisers and advertising agencies, as well as celebrities in the economic sub-strata concerned with exploitation of the claim on personality. In an inflationary economy, advertising costs can, within surprising limits, be fully passed on to consumers of the goods being advertised. This is even more likely to be so in the case of mass marketed products wherein the cost of advertising forms a relatively small percentage of the cost of sales. Moreover, armed with an armada of data which can justify celebrity endorsements and indeed virtually require them on a cost-benefit basis agencies have little trouble convincing their corporate clients to pay six or even seven-figure fee for major endorsements.

Moreover, advertising agencies often have no direct interest in keeping media costs as a whole low. There is a well understood, if not easily comprehensible, tradition in the advertising business, that an agency gets paid on a percentage basis of media costs which will often include the fees of endorsers. Naturally, this is subject to negotiation in particular cases, but advertisers must indeed have a great deal of clout to negotiate with major agencies. I am not for a moment suggesting that advertising agencies act improperly in this respect: the

---

1 Treece, Commercial Exploitation of Names, Likenesses and Personal Histories, 51 Texas Law Review, 637 at 644
rules of the game are widely understood and would be corrected by the free market if there were any intense desire to do so. However there does not appear to be such a predilection at the present time.

Moreover, in addition to the foregoing, there are intangible rents prevalent in the celebrity advertisement phenomenon which are undoubtedly influential. Which of us, if given responsibility for an advertising campaign, would not somewhat cherish the thought of having cocktails with Brooke Shields or Candice Bergen?

Given that celebrity endorsement advertising tends to work in the viewpoint of advertisers, and given the huge fees that celebrities receive, it would appear prima facia to be the case that rents exist among all players in this game, and that the rents are, ultimately of course, borne by the consumers of the products and services involved. More on this point and the subject of elasticity will follow later.

Returning to Globerman's pattern of analysis, let us look at the first parallel situation:

1. In codifying by law (whether common law or statute law) wherein the claim on personality already virtually exists under Canadian law and along the lines of the more conservative American jurisdictions, there is unlikely to be any significant change in the economics of the system. What may happen, however, is that increased awareness of the law on the part of some of the major economic players could result in a reallocation of existing rents. For example, a would-be celebrity endorser in Canada now more sure of his legal footing and not limited to the three figure award given in the Athans case for abuse of this right, might begin to demand much higher fees. Likewise, certain American interests which have hitherto received more or less nominal recognition of the right of publicity in Canada due to vagueness under our laws might extract higher fees in some cases and lower fees in others (i.e. if it were clear that the right was non-descendible and the personality involved were dead).

2. Concerning the creation of a specific right where a general one exists, or in the instant case attempting to codify absolutely precise ambits of the right and to possibly set minimum payment requirements, this is probably a parallel
that would have no application in the present context. The only possible application of this particular scenario of Globerman\(^1\) would be in setting by law some sort of minimum scale of payment. In the case of professional actors, this is already done by means of SAG rates in the U.S. and ACTRA rates in Canada, which are for services and not for endorsements as such. In the case of non-professional persons dealing with the right of publicity for purposes of commercial endorsement, they will inevitably have to deal either with SAG or ACTRA if only on a work-permit basis and will receive at least the minimum fees. In any case, the going rate for top-calibre celebrities may well be several thousand times the minimum "union" rates.

3. Globerman's third scenario\(^2\) deals with the creation of a re-use right where one does not exist. Our parallel would be the creation of a right of publicity where one does not presently exist, except perhaps in certain more venturous American jurisdictions tending towards recognition of the descendibility theory of the right of publicity. Part of Globerman's analysis of this scenario\(^3\) is confined to the context of constrained bargaining involving unions. This would almost certainly not be the case, as SAG and ACTRA have shown a pronounced and thoroughly laudable hands-off approach to the situations of celebrities and superstars, and these are likely to be the only persons normally affected by a descendible right of publicity.

In their other sub-species of this category, the Globerman authors deal with the free bargaining situation which would be applicable here. They conclude in this regard that "if there are rents in the system, it is at least possible that returns to performers can be increased. The entire result depends on the current distribution of rents."\(^4\)

If I am correct about the existence of rents in the advertising sector and in the hands of celebrity endorsers, then almost certainly the income of many holders of the alleged "right of publicity" could demand even higher fees for the right.

---

1 Ibid, Page 60
4 Ibid
while they were alive because the right would have a useful duration even after they die.

This is not so clear in endorsement situations which are the most dramatic and remunerative form of exercise of the right of publicity. This is simply a matter of advertising tastes. While the public may become very excited to know that Bill Cosby likes Jello, they will not be very interested in this fact after he is dead and would probably be somewhat appalled to have such an endorsement thrust upon them. One occasionally sees such advertisements for certain very highbrow products such as cigars or spirits, and a vogue may yet develop for such post-mortem endorsements in wider areas. However, the primary market for exploitation by celebrities motivated by estate planning concerns would likely be in the area of merchandising.

Merchandising of the celebrity himself, i.e. T-shirts, statues, buttons, posters, games, sports cards, hockey sticks, golf balls and whatever other printed, mechanical or even electronic paraphernalia can be devised, will certainly become more expensive because would-be purveyors of these items will no longer find it necessarily financially advantageous to wait until the death of the celebrity in order to deal with these items on a royalty-free basis. The celebrity will likely enter into extensive licensing agreements during his life knowing full well that a right is enforceable for a distinct period (often advocated as fifty years) after his death. Whether or not the value of these rights is capitalized largely at the beginning, or largely left to be paid for as royalties to an estate or assignee corporation, would depend on a host of factors including the celebrity's life expectancy, his particular type of status, the particular merchandise involved and many other matters.

If such rights could, in effect, be exercised from the grave, the public will be forced to pay a significant premium for informational and cultural content to which they have become accustomed to receiving essentially as "public domain".

Much of this content would not be really disseminated under the rubric of first amendment freedom or the corresponding "newsworthiness" doctrine under
Canadian law. It would be a significant disruption in the marketplace and a significant aggrandisement to the estates of a few already incredibly wealthy celebrities.

It would furthermore create even more distortion to the elasticity concept which is already highly strained in analysing the incomes of superstars, particularly in the field of endorsements. While there exists a substantial number of trained actors, actresses and celebrities who are perfectly willing to do endorsements, a sponsor may decide that there is only one such person whose image fits. A sponsor, working from this premise; may be willing within non-apparent reason to pay almost anything for this endorsement. Obviously at a certain point the price quoted by the endorser would become prohibitive and the sponsor would start looking elsewhere. However the six and seven-figure prices currently paid for celebrity endorsements in some instances certainly appear to defy the normal economic rules by which most of us set the price for goods and services we supply. After all, in substance it is usually only a few hours of work and time by the celebrity that are specifically required to create a commercial. The "rent" received by the celebrity is, as we have seen, paid for by the consumer of the product that is being endorsed.

I am not suggesting for a moment that there is anything legally or economically wrong with the present legal status quo which enables celebrities to make these phenomenal incomes. In the absence of a definitive study on the matter, I suspect that economic distortions caused by such anomalies as measured on a macro basis are relatively minimal. However in more localized terms, they can be and probably are fairly significant. This would certainly be the case if a person other than a celebrity or an oil sheik tries to buy real estate, or almost anything else for that matter, in Beverley Hills.

Elasticity, in this context is defined by Globerman as "the proportionate relationship between the change and the quantity of performers' services offered on the market and the change in real wage rate offered for such services, other things constant." ¹

¹ Globerman, Page 36
I doubt that there is any "shortage" of superstars and celebrities available. The only situation with which we are concerned at the superstar level is whether a particular person is willing to offer his services in a commercial way, for a particular price. If that price goes up substantially as it inevitably would if rights become descendible, certain persons who may hitherto have been unwilling to endorse certain products may come forward.

Clearly, at a certain point the public will be starting to pay tangibly for media portrayal of "information" (I use the word in a deliberately non-committal sense as encompassing both advertised factual matters on the one hand and clearly news material or first amendment protected material on the other hand) to an extent that it never has before. Wealth would be created in certain elite circles for services which were never really performed and in all likelihood on the basis of work done and careers devised which would have all happened anyway without the additional career incentive. One doubts that even the most avid advocates of the career incentive theory would say that Johnny Carson is underpaid or that his artistic and creative legacy to mankind in general would increase because he thought the specific legacies to his heirs would be larger than they already will be.

One might also attempt to argue that as the cost of procuring the right to use established celebrities' personalities becomes too high, the media will do what they have shown themselves capable of doing many times in the past, namely inventing or creating new celebrities at a cheaper price. However the entertainment industry is not nearly so rational in this respect as other sectors of the economy. In general, once a star is created, that star can demand almost ethereal compensation as long as his product is saleable. Marlon Brando now reputedly receives several million dollars for brief "cameo" appearances in such movies as "Apocalypse Now" or "Superman". One cannot help but wonder whether the decision to pay these kinds of fees was based on a reasoned assumption that the cost would be met by an even greater benefit in
terms of box office sales, or whether in fact the box office sales were likely to be so
great in any event that the more intangible and highly rent-prone subjective factors at
work in Hollywood led to such types of contracts.

Even the courts which are tending in the direction of descendibility
recognize or decree that the right which gives rise to the descendibility theory must
be exercised during the celebrity's lifetime. This can be seen as an attempt to limit
the "flood gate" implications of creating a new property right which is arguably very
expensive to the public and which may have an almost open-ended duration. However from
an economic standpoint even this distinction seems relatively meaningless. If I own
a piece of land during my lifetime but have not developed it, is it any the less
devisable to my heirs? If I do develop it during my lifetime it has more value to my
heirs, but only because I have spent time and money on it. Likewise, certain distinctions
are made under copyright law regarding works which were published or not published
during an author's lifetime. However the distinction has implications only to the
duration of copyright, not to the existence of it.

Either we are dealing with a property right, or we are not. To
create a property right and, on the bootstraps of its creation, make it descendible, is
only going to aggrandize the heirs of a smaller number of extremely wealthy individuals.
even further. The public will very likely pay for it in a tangible way every time they
purchase a service or a product that has ever been sponsored or every time they purchase
certain information, the content which they now receive essentially for free. This,
in my view, would be precisely the kind of situation envisaged by Mr. Justice Estey
in Krouse v. Chrysler\(^1\) in which he states "the danger of extending the law of torts to
cover every such exposure in public not expressly authorized is obvious. Progress in
the law is not served by the recognition of a right which, while helpful to some
persons or classes of persons, turns out to be unreasonable disruption to the community
at large and to the conduct of its commerce."

\(^1\) 1 O.R. 2nd 225 at 240
CONCLUSION

The writer has come to a conclusion which is somewhat embarrassing in the context of a conference dedicated to economic analysis of intellectual property rights. The conclusion is that the so-called "right of publicity" is not really a property right in any known nature of personality at all, and is indeed probably a perfidious misnomer. It is, or should be treated as, a cause of action sounding in relatively traditional doctrines of tort law, with perhaps a mild dosage of quasi-contract or restitution remedies thrown in to fill in the gaps caused by assessment of damages problems in traditional tort law. Even this latter digression may be largely unnecessary if the courts avail themselves to a greater extent of the doctrine of exemplary or punitive damages. In order to mollify partisans in the property vs. tort debate, the writer is prepared to suggest that the phenomenon be renamed that of "claim on personality".

© HOWARD P. KNOPF, 1983

Howard P. Knopf is a Toronto lawyer practising in the entertainment, media and communications fields.
International Aspects of Copyright Legislation in Canada: Economic Analysis of Policy Options

John Whalley
University of Western Ontario

November 1982
I Introduction

This paper discusses some of the issues which arise in deciding on an appropriate policy stance for Canada on the question of copyright protection for foreigners. Canada's present international copyright obligations are governed by the 1928 Rome text of the Berne Convention and the 1952 text of the Universal Copyright Convention (UCC). Both of these involve 'national treatment' for foreign holders of copyright, with the UCC by far the most important since it, rather than the Berne Convention, applies to Canadian copyright holders whose works are sold or used in the U.S., and U.S. copyright holders whose works are used or sold in Canada.

The main options for Canada, as outlined in the Keyes-Brunet (1977) Report, are: (i) Accession to later texts of either or both of these conventions; (ii) withdrawal from either or both; and (iii) maintenance of the status quo. At first sight, the appropriate direction for policy toward copyright protection for foreign works in Canada would seem to hinge on the size of flows of copyright royalties in and out of Canada. If Canada is a net payer of royalties abroad, then by not protecting foreign works through copyright Canada would seem to be made better off by withdrawing from existing international arrangements. This would be the case even if foreign countries retaliate by not protecting Canadian works abroad, since on a net basis a saving of cash occurs which, in turn, would seem to be the net gain.

In practice, however, things are more complex. This paper argues that it is inappropriate to consider international copyright protection in isolation from other international arrangements. Treaty agreements exist not only in the copyright field but in trade, tax, and other areas. Abrogation of international copyright agreements can thus provoke retaliation in other areas. The key question is whether current non copyright treaties already represent optimal arrangements for foreigners. If so, they have no incentive to change non-copyright treaties using unilateral action by Canada on copyright withdrawal as a pretext. However, given the dominant importance of the U.S. trading relation-
ship to Canada, and the fact that the main treaties on trade protection are multilateral rather than unilateral such a position, to the present author at least, seems naive.

This paper also argues that alternative policy approaches are available for Canada toward foreign holders of copyright besides those outlined by Keyes and Brunet. While existing copyright conventions discuss at some length the treatment of foreign copyright holders under a non-discrimination national treatment principle, no reference appears to be made to tax treatment of copyright royalties. Canada currently has a zero withholding tax rate on copyright royalties being paid abroad. Canada would appear to be able to increase this withholding tax rate without violating the principles of either the Berne Convention or the UCC. A penal tax treatment of royalties paid abroad is, of course, close to withdrawal from existing treaties, but does not raise the same legal questions of abrogation.

A further area of alternative policy approaches for Canada concerns those areas of ambiguity of coverage in existing treaties. These issues are touched on by Keyes and Brunet and cover performers rights, library lending, computer programmes; and many other areas not covered by the texts to which Canada has thus far acceded. A 'hard' stance on these issues by Canada limits the extension of copyright protection for foreigners into these new areas and now applies to a significant enough portion of the whole area to make it a question worthwhile for policy makers to consider.

Finally, this paper argues that while it may seem straightforward to assess gain or loss to Canada from participation in or withdrawal from existing copyright conventions by measuring the net royalty flow, this is in fact not the case. Copyright protects monopoly rents which also have to enter the calculation. For artistic works whose revenues at point of production are uncertain, the ex-post monopoly rent which copyright protects maybe a further significant
source of saving to Canada by withdrawal from or modification to existing copyright agreements. Quantification using the net royalty flow alone is therefore treacherous.

II International Copyright Protection in Canada

The position of foreign holders of copyright to works used or performed in Canada is governed by two separate international conventions to which Canada is a party. The first is the Berne Convention which dates from 1876 with the 1928 Rome text containing the applicable provisions for Canada. Later texts exist to which Canada has not acceded. Because of the non-participation of certain countries (most notably the U.S.) in the Berne Convention, a further Convention (the Universal Copyright Convention (UCC)) links Berne and non-Berne countries. Canadian participation is based on the 1952 text, even though a later 1976 text exists to which Canada has not acceded.¹

In simple terms copyright protection in Canada applies to works of authorship and creative ideas expressed through performing arts and other media. Currently, copyright protection gives the copyright holder the right to receive royalties during his lifetime plus an additional 50 year period. Under the Berne and UCC Conventions signatures to these conventions agree to extend copyright protection to foreign works as well as those by domestic residents. The basic principle is one of 'national treatment' with each member country providing nationals of other countries the protection accorded under their own domestic law.

The international treatment of copyright has been somewhat complicated over the years by the different structure of copyright prevailing in the U.S. The U.S., for many years maintained, a complex procedure involving notice, registration, deposit, affidavit, and manufacturing requirements. Under the domestic manufacturing provision, copyright was only given to foreigners if manufacture of the work involved took place in the U.S. The Universal Copyright Convention

¹More detail on these arrangements is given in Keyes and Brunet (1977), and Torno (1981).
(UCC) adopted in 1952, reflects the attempt to find a common denominator between the provisions of the Berne Convention and the copyright structure in the U.S. Under this convention copyright provisions applied to foreigners apply equally to Berne and non-Berne countries, the most important which are the U.S., the Soviet Union and China. The compromise in the UCC is to significantly moderate the requirements for notice and registration, which are now met by the appearance of the copyright mark © along with the name of the copyright claimant and location. Through the UCC Canada is thus exempt from the prohibitive provisions of U.S. copyright practice.

As already mentioned, this copyright protection makes no mention of tax treatment of royalties. It simply gives the holder of copyright the ability to control the use of his or her work during the period in which the copyright protection applies. A further feature of these conventions is that they contain no exhaustive classification of copyright. While there was general discussion at the time the texts were being drafted of the extent of which copyright protection applies no clear guidelines are given as to how certain specific issues are to be treated. These involve the question of performing rights for sound recordings, whether or not artists should receive copyright payment every time their works are performed rather than simply when sold, whether with home recording the copyright holder of the item being recorded should qualify for a royalty payment, whether copyright holders should receive royalties for record rentals, lending, and all photocopying.

All of these areas are subjects for discussion at the present time as to how far the extension of copyright in Canada should go. The question of extension in these areas to foreigners forms part of the set of questions arising with the international treatment of copyright. In addition to these ambiguous areas, there are also new products which arise from time to time where
copyright protection is not clearly established. This is the case, for instance, with cable companies where the retransmission of TV and radio signals did not involve paying a royalty to the copyright holder. The only permission in Canada which is required is from the CRTC. Home recordings and record rentals also fall into this area simply because these were not envisaged as contentious areas of copyright when the international treaties were initially negotiated.

Thus the whole area of copyright for Canada involves first of all whether Canada should abide by the terms of the treaties into which she has entered in earlier years, how these treaties should be interpreted to apply to particular products, and how extensions of those treaties should or should not be undertaken to cover new products and areas reflecting subsequent developments after the initial signing of the treaty.

III Options for Canada

In approaching the question of international copyright protection a number of different policy options and issues arise for Canada. These can be listed as: (a) withdrawal from existing copyright agreements, (b) changes in tax treatment of copyright royalties, (c) unilateral moves to renegotiate treaty provisions, (d) issues concerning the extension of existing copyright law to new products and new areas and, (e) partial withdrawal from copyright provisions, such as changing the length of the period of copyright protection under current arrangements.

(a) Withdrawal

An issue taken up in the Keyes-Brunnet report is whether Canada is made better off or worse off by the structure of international treaties under the Berne Convention and UCC. Does Canada gain or lose by participating in this multilateral framework of treaty agreements which mutually protect copyright holders in all countries?
At first blush, a simple analysis of this question would suggest examining whether or not Canada gains or loses from the flows of royalty payments in and out of the country. Royalties paid abroad represent payments to foreign holders of copyright for copyright use in Canada, and royalty payments received from abroad represent payments to Canadians from the use of copyrighted materials abroad where the copyright holder resides in Canada. If a positive flow is involved for Canada, Canada would appear to be a net gainer from international copyright arrangements. If the net flow is negative, Canada is a loser.

Unfortunately this question is not simple as it seems at first sight for a number of reasons. Firstly, if Canada were to withdraw from existing copyright agreements and treaties this has wider implications beyond the copyright area. Canada is a signatory of many international agreements including the General Agreement on Tariff and Trade, and bilateral agreements such as the international tax treaties in which Canada has entered. Were Canada to withdraw from the copyright agreements it raises the issue of precedent for unilateral abrogation of a treaty and, in turn, potential retaliation by trading partners in other areas outside of copyright which might be harmful to Canada.

The perception of Canada in the U.S. at the present time would seem to be that of a relatively small country which is determined to flex its muscles with regard to economic nationalism. Pressures appear to be building in the U.S. Congress for retaliatory moves of various kinds if Canada continues on or accelerates her present course. These pressures are reflected in recent U.S. regulations concerning cross border trucking, and also the ongoing controversy with respect to tax treatment of conventions and advertising in the U.S. by Canadian corporations and the threats in the Congress to retaliate through
compensating tax measures there. To the extent that existing arrangements with the U.S. do not represent optimal U.S. policy, a clear danger for Canada with withdrawal from copyright arrangements is that this will be seen as a part of a set of restrictive measures in Canada which may, in turn, provoke retaliatory action in the U.S. with the associated losses involved for Canada. The quantitative importance of copyright issues to Canada would seem to be too small to allow this issue to be the straw that breaks the camel’s back and encourages an elevated protectionist stance against Canada in the U.S.

A second issue concerning potential withdrawal from international copyright agreements by Canada is the question of how large are the gains that Canada would reap from withdrawal or, in turn, how large the losses are from current arrangements. A number of reports on copyright in Canada have attempted to quantify the size of the royalty flows in and out of Canada, but unfortunately the data available are not of high quality and precise indications as to the royalty flows involved are not easily obtainable. Keyes and Brunet examine data on trade flows in copyright materials, but as they say this gives only an imprecise indication as to the bilateral flows of copyright royalties themselves. One of the most useful pieces of evidence, however, is that from the Statistics Canada CALURA file of consolidated corporated accounts. From this file it is possible to determine what portion of copyright royalties paid in Canada go abroad, and to which countries. Data for the early 1970’s indicate (Keyes and Brunet, p. 32) that only about 15% of all royalties originating in Canada remain in Canada; 85% are paid abroad with the overwhelming majority (95-97%) being paid to the U.S. While this does not provide information on copyright royalties received from abroad, the size of the flows paid abroad is strongly suggestive of the position that Canada is a significant net loser from copyright royalties paid abroad. In 1973 these payments were in the region of $42 million.
The size of the net royalty flow abroad, however, is not the only consideration in evaluating Canadian gain or loss. One of the most important characteristics of works of authorship is the revenue uncertainty involved at the point of production. This applies equally to authorship and recording activity. When an author writes a book he or she has only an expectation as to what his eventual royalties and sales may be and it is only after marketing that it is known what the sales of his work actually are. Thus in terms of the copyright royalties being paid abroad from Canada, the potential saving to Canada is not only the royalties themselves but also the ex-post monopoly rents which are protected by copyright provisions. Were Canada to unilaterally withdraw from existing copyright arrangements, these monopoly rents would no longer need to be paid to the holders of copyright located abroad. This suggests the potential gains to Canada from withdrawal from copyright arrangements could be larger than the savings in payments of royalties abroad, although foregone monopoly rents received from abroad also need to be added in. No data currently seem to be available to accurately measure these rents.

(b) Tax Treatment

Existing copyright conventions make no mention of the tax treatment of copyright royalties paid abroad. Currently in Canada there is no withholding tax on copyright royalties paid abroad, and a way for Canada to harshen its treatment of foreign copyright holders (effectively partially renegotiating copyright arrangements) would be to introduce a special withholding tax on copyright royalties including different rates on copyright royalties paid abroad. There would undoubtedly be questions as to whether or not such tax treatment would violate the U.S.-Canadian double taxation treaty or other tax treaties Canada has entered into. Since these tax treaties typically specify withholding tax rates on patent royalties rather than copyright royalties there
would seem to be no abrogation of treaty involved.

This also raises the question of what would be the optimal tax treatment for such royalties. On the one hand, if Canada is a small open price taking economy then it would seem there are no disincentive effects by having penal tax treatment. In this case the flow of copyright materials is no way affected by the high taxes which Canada charges. In effect Canada can free ride on the creative works provided by foreigners. On the other hand, to the extent to which the flow of original works from abroad into Canada is slowed by this tax treatment then some optimal tax treatment would seem appropriate.

A further set of questions arises with the tax treatment of copyright royalties elsewhere, and the extent to which the tax treatment of royalties abroad is equal to that in Canada. While the U.S. has a zero withholding tax rate on copyright royalties, it would appear that a number of European countries have withholding tax rates which are non zero, raising the question of whether or not Canada would be better off to follow this treatment.

(c) Unilateral Attempts at Renegotiation

In addition to the possibility of withdrawal from existing treaty obligations, a further option for Canada is to seek unilateral renegotiation of features of these treaties. Rather than withdrawal, under this strategy Canada would make proposals for realignment of these treaties which were in Canada's interest such as a reduction in the period of time for which copyright applies.

As a policy option, this approach seems unlikely to succeed. The number of countries involved in both the Berne Convention and the UCC is now in the order of 60 to 70 and unilateral moves for renegotiation with such a large group of countries is almost certain not to succeed. It has proved difficult in the past to get agreement on renegotiated texts for the Berne Convention and
the UCC. Thus, any attempt by Canada to move on unilateral renegotiation is likely to be successful.

(d) Other Areas and Extensions of Copyright

As already mentioned, there are currently a range of issues on the agenda for possible extensions of copyright each of which has international implications. These include: performing rights for sound recordings and the payment of royalties on these, royalty payments for performances, a renegotiation of the compulsory licensing arrangements for sound recordings, issue concerning home recording and whether or not a blank tape levy or tax should be charged, the wider issues of record piracy and counterfeit, record rentals and the right of performers and composers to receive a royalty, royalties to authors from library lending, and payment of copyright to accompany photocopying.

All of these are areas where attempts are under way to extend copyright by broadening existing copyright protection. In all these areas the international aspects of possible extension come into play, since foreign holders of copyright typically benefit from these extensions. To some extent, then, this gives Canada an opportunity to change its international copyright treatment relative to other countries by not agreeing to these extensions, since no abrogation of treaties involved and the savings may be substantial.

IV Wider International Aspects of Copyright Law

In addition to the issues raised concerning policy options for Canada on international treatment of copyright, at the end of the day the key issues may well be the wider international aspects of any possible international copyright renegotiation. Currently, Canada is involved in a series of international negotiations on the GATT, UNCTAD, North-South forums, tax and trans-border pollution, and many more issues. The question of changing Canada's treatment of foreign holders of copyright is thus part of a wider set of policies which may or may not represent an attempt to assert a degree of Canadian nationalism.
Within the GATT a sequence of negotiating rounds have been under way since 1947 with the most recent of these, the Tokyo Round, establishing a set of codes for certain selective non-tariff barriers and also tariff cuts which are to be instituted by 1987. Already there have been major disagreements among trading partners within the GATT as to the effect of these codes and the GATT ministerial scheduled to start in November 1982 is designed to take up some of these grievances. In addition a number of trade frictions are currently taking place involving Canada. In one case, the issue of FIRA, a reference has been made by the U.S. to a GATT panel who are to rule on the question of whether or not any violation of GATT Agreements is involved through the operation of FIRA and their trade content provisions before investment in Canada is authorized.

All of these frictions are evidence of a growing international climate of tension between Canada and her trading partners, including the EEC and Japan. The view which is argued for by many commentators on this wider situation is that Canada as a smaller country gains from the multilateral approach to world trade liberalization which has characterized the world since the second world war. Anything which jeopardizes this multilateral approach in which the smaller countries gain is thus disadvantageous to Canada, since there are a range of opportunities which exist for the larger countries to subvert the effects which have been produced by multilateral trade liberalization. As an example, it is relatively simple for the U.S. to modify the codes on government procurement which have been instituted so as to effectively undo their containment in the Tokyo Round. In fact, many commentators in Canada argue that the major benefits to Canada from participation in the GATT Tokyo Round was the increased access to the government segment of the market place in the U.S. through a liberalized government procurement code.
Thus, although the issues narrowly in the copyright area seem to indicate that Canada can potentially make themselves better off by withdrawal from the current set of international conventions, because of the dominance of one trading partner for Canada and the signal which would be created by such withdrawal it would seem at the present time that such policy action is probably misguided. Indeed one can argue conversely that for Canada to move closer to the U.S. position on copyright might be advantageous because the quantitative orders of magnitude involved may well be quite small, and if this maintains existing arrangements elsewhere in the international trade area then the gains to Canada are probably highly significant. The wider aspects of trade policy thus crucially come into play in any approach to redesigning copyright arrangements in Canada.

V. Conclusions and An Evaluation - Options for Canada

This paper argues that considered narrowly, Canada is almost certainly a loser from the current international agreements she has with respect to copyright. Although the data currently available are imprecise, the indication is strong that Canada is a substantial net payer of copyright royalties abroad. This net flow, it is argued, probably underestimates the gain to Canada from withdrawal from existing treaties since the monopoly rents protected ex post by copyright have to be added in.

The narrow considerations, however, are deceptive since compared to other international trade and investment policy issues, copyright issues appear quantitatively insignificant. If one's view is that Canada as a smaller country is a gainer from current multilateral trade liberalization arrangements (such as under the GATT), it seems misplaced to suggest policy changes which may provoke or contribute to a climate which may provide retaliation on the
the wider front of tax and trade questions. The key question is thus are policy arrangements in non-copyright fields already optimal for foreigners so that action on the copyright front will provoke no wider response? Or are these other arrangements governed by fragile multilateral agreements with the larger countries (such as the U.S.) potentially capable of being provoked into a retaliatory response by a precedent-setting action on copyright. While this is undoubtedly a tough policy call to make, this author's instinct is towards the latter position.

VII. References

Keyes A. A. and C. Brunet [1977]. Copyright in Canada: Proposals for a Revision of the Law, Consumer and Corporate Affairs, Canada.

Torno, B. [1981]. Ownership of Copyright in Canada, Consumer and Corporate Affairs, Canada.