

The Economic Impact of Patents in a Knowledge-Based Market Economy**

by Bojan Pretnar*

NOTE: THIS PAPER WAS ORIGINALLY PUBLISHED IN INTERNATIONAL REVIEW OF INTELLECTUAL PROPERTY AND COMPETITION LAW (IIC) VOL.34 NO. 8/2003, PP. 887-906.

I. Introduction

There is little doubt that the mainstream view in economics sees patents as a tool for granting a monopoly to patentees.¹ Apart from noting numerous single-prize, winner-takes-it-all models of patent races that dominate in economic literature, this view can easily be verified by opening almost any textbook on economics or industrial organization. The monopoly doctrine of patents is usually presented on the basis of the seminal paper by Arrow,² which actually does not deal explicitly with patents. His analysis, limited to purely cost-reducing innovations, is frequently presented in a diagrammatic version developed by Demsetz,³ which has also been used to explain the basic logic of patents for product innovations, for example by Scherer and Ross⁴ and more recently by Maskus.⁵

On a more advanced level, one can of course find a number of models in which patents are analyzed in a competitive setting, though usually within a more specific framework.⁶ However, this fact does not change the basic truth that the “patent-equals-monopoly” doctrine dominates the field.

On the other hand, a number of arguments cast considerable doubt on the appropriateness of the mainstream theory.⁷ Most importantly, the bulk of empirical

** The author is grateful to Dr. Branko Ilič, Faculty of Social Sciences of University of Ljubljana, for his invaluable contributions, and to Prof. Keith E. Maskus, University of Colorado, Boulder, for comments on an earlier draft of the paper. The views expressed by the author in this paper are strictly personal and do not necessarily reflect the views of the World Intellectual Property Organization (WIPO), of which the author is currently a staff member.

* Professor, Faculty of Economics at the University of Ljubljana, Slovenia.

¹ Cf. EDMUND W. KITCH, “Patents: Monopolies or Property Rights? Research in Law and Economics” 31-41 (Vol. 8, 1986).

² KENNETH J. ARROW, “Economic Welfare and the Allocation of Resources for Invention,” in: “The Rate and Direction of Inventive Activity: Economic and Social Factors” 609-626 (National Bureau of Economic Research, Princeton University Press, Princeton 1962).

³ HAROLD DEMSETZ, “Information and Efficiency: Another Viewpoint,” 12 Journal of Law and Economics 1-22 (1969).

⁴ FREDERICK M. SCHERER & DAVID ROSS, “Industrial Market Structure and Economic Performance” 622 *et seq.* (Houghton Mifflin Company, Boston 1990).

⁵ KEITH E. MASKUS, “Intellectual Property Rights in the Global Economy” 29 *et seq.* (Institute for International Economics, Washington, D.C., August 2000).

⁶ See, e.g., MICHAEL WATERSON, “The Economics of Product Patents,” 80 AER No. 4, September 1990, at 860-869.

⁷ KITCH, *supra* note 1, at 31 *et seq.*

research undertaken in the last decades repeatedly suggests that patents can hardly be considered as monopolies.⁸ Last but not least, patent statistics speak a plain language. One may choose any particular product and then search for related patents; in most cases, hundreds, perhaps thousands, of them would be found.

The notorious discrepancy between mainstream theory and empirical evidence is not only paradoxical, it is a matter of concern. The economic importance of patents is rapidly increasing. Thurow, for example, believes that the knowledge-based economy will not even work without what he calls “a new system of intellectual property rights.”⁹ Whether such a new system is indeed necessary, and if so, what its framework might be are questions, which cannot be adequately addressed unless a sufficiently precise and unambiguous understanding of how the *current* patent system works is available. Yet the disappointing fact is that the mainstream economic theory is far from being convincing .

The aim of this paper is, therefore, to identify the sources of persisting discrepancies and seek an alternative explanation of the principal functioning of the patent system, which would better withstand the test of the reality of a modern, knowledge-based economy in which innovation-based competition, not monopoly, prevails.

II. Pitfalls and Misconceptions of Prevailing Economic Doctrine

It may be useful to begin with the well-known fact that the mainstream doctrine is based on Schumpeter’s postulate that monopoly profits constitute the economic incentive to invent. More precisely, monopoly profits are assumed to be *the only* reward for undertaking costly and risky inventive activity, because competitive pricing “generates no rents with which to cover the costs of the original research and development program.”¹⁰ It is then no wonder that Arrow explicitly equates incentive to invent with the potential monopoly profits achieved by an invention.¹¹

Given the peculiar nature of inventive activity, the output of which is information as an inherently public good, patents are inevitably interpreted to be the means that make the innovative reward possible. This logic has been widely accepted as the classic function of the patent system ever since 15th century, when patents were first granted in Venice.¹² As a consequence, patents are traditionally interpreted as a monopoly, which is thus a kind of trade-off between static and dynamic efficiency.¹³ However, it must be realized that such an interpretation is *not* a logical conclusion derived from the model, but rather an inseparable part of the core assumption. If patents are not interpreted as monopolies, then the assumption of rewarding innovation with monopoly profits simply cannot be maintained.

⁸ Cf., EDWIN MANSFIELD ET AL., “Imitation Costs and Patents: An Empirical Study,” 91 *Economic Journal*, December 1981, at 907-918; RICHARD C. LEVIN ET AL., “Appropriating the Returns from Industrial Research and Development,” *Brookings Papers on Economic Activity* No. 3, 1987; I. COCKBURN & R. HENDERSON, “Racing to Invest? The Dynamics of Competition in Ethical Drug Discovery,” 3 *Journal of Economics & Management Strategy* (3), at 481-519.

⁹ LESTER THUROW, “Creating Wealth” 125 (Nicholas Bailey Publishing, London 1999).

¹⁰ MASKUS, *supra* note 5.

¹¹ ARROW, *supra* note 2, at 619.

¹² SCHERER & ROSS, *supra* note 4, at 621.

¹³ Cf. MASKUS, *supra* note 5, at 28 *et seq.*

Despite being almost indisputedly accepted, the above logic can easily be identified as the first – and crucial - source of inadequate interpretation of patents. The reason is simple: Schumpeter’s assumption of “monopolistic” incentive to invent is merely a behavioral assumption but not a law of economics, as pointedly remarked by Liebowitz.¹⁴ Pursuing innovation in quest of monopoly profits was possibly a reasonable assumption in the past, but cannot be taken as a representative paradigm in the context of a modern, knowledge-based economy. The modern world is characterized, *inter alia*, by what Baumol calls routinization of inventive activities,¹⁵ the consequence of which is a competitive outcome, not a monopoly:

The research and development laboratories of two competing firms often turn out innovations that are imperfect substitutes for one another. One may be superior to another in some features but inferior in others. Or one may be only marginally inferior and therefore salable at a price that is a bit lower than the other’s. In such cases, as already suggested, the race offers many prizes; the winner receives the highest payoff, but those who come in close behind obtain compensation commensurate with the value of their performances.¹⁶

In addition, the way in which patent law itself is interpreted in economics is also having a considerable impact on the mainstream doctrine. The scope of patent protection is regularly assumed to be much broader than actually determined by law. It is then no wonder that the assumption that intellectual property rights in general and patents in particular confer an economic monopoly on their owner is considered by Kitch to be one of the “elementary but persistently repeated errors” in economic analysis.¹⁷ Such a state of affairs thus calls for a closer look into the main reasons of noted errors and observed discrepancy.

The first and perhaps most essential reason is the persistent lack of a systematic treatment of patents as an unusual economic construct: *the law has made patents to be simultaneously both private and public goods*. As a legal right, a patent makes an invention a typically private, *i.e.*, scarce, economic good, but only insofar as the patented invention cannot be exploited by a third party in products or related processes without the consent of the patent owner. During the term of their validity, patents thus prevent the use of “protected” information only with regard to its *embodiment*, either by patentees themselves or their licensees, in related products. This is the *private-good effect* of a patent, which affects Schumpeter’s innovation phase of the whole innovation process.

On the other hand, patent law has always required that all the information contained in (valid) patents must be fully disclosed. This absolute requirement clearly has an important impact on Schumpeter’s diffusion phase. More important, however, is the fact that the disclosed information is available *as a free, costless input into the production of new knowledge at any point of time, that is, also during the period of validity of patents*. Note that production of new knowledge *as such* is a distinct process from that of embodying a certain invention in a product. Production of

¹⁴ STAN LIEBOWITZ, “Re-Thinking the Network Economy” 3 (Amacom, New York etc. 2002).

¹⁵ WILLIAM J. BAUMOL, “The Free-Market Innovation Machine” 4 (Princeton University Press, Princeton 2002)

¹⁶ BAUMOL, *supra* note 15, at 28.

¹⁷ EDMUND W. KITCH, “Elementary and Persistent Errors in the Economic Analysis of Intellectual Property,” 53 Vanderbilt Law Review, November 2000, at 1729.

knowledge is characterized by the fact that, apart from human labor and capital, both the main input (stock of available knowledge) and output (new knowledge) are intangible. Such newly generated knowledge then serves in turn as the core input in another process of transforming an invention into an embodied innovation, the output of which, however, is now a tangible product.

In other words, information contained in patents is always a pure public good with zero price whenever used for the particular purpose of generating new knowledge, that is, in Schumpeter's invention phase. This is then the *public-good effect* of a patent, by which cumulative research is explicitly allowed, indeed promoted, so that innovators can stand on the shoulders of the giants who precede them, as Scotchmer would say.¹⁸ The public-good effect of patents then obviously opens the way for innovation-based competition to take place, provided, however, that the pro-competitive nature of the public-good effect is not subsequently impaired by the private-good effect.

It is somehow strange to note that this double economic nature of patents, despite being regularly noted, has seldom been analyzed as a whole in a mutually consistent manner. After all, the origin of the word "patent" itself ("open letter") and the well-known rationale of the social contract theory of patents is built precisely on their simultaneous private-good and public-good effects.¹⁹ On the other hand, it seems that those not trained in patent law may easily be misled by legal jargon. For example, U.S. law provides that the owner of the patent has the right to exclude others from making, *using*, or selling the patented invention²⁰ (emphasis added). The scope of the basic right to exclude others from using the patented invention must be carefully construed for purposes of economic analysis, because the free use of information disclosed in patents for purposes of producing new knowledge (the public-good effect) is *not* – and has never been – prevented by law. Yet this fact is often dealt with in an ambiguous manner in the field of economics. The following two quotations, coming from two Nobel laureates in economics, may help to illustrate the point.

In his famous paper, Arrow states, *inter alia*, the following:

In the absence of special legal protection, the owner cannot, however, simply sell information on the open market. Any one purchaser can destroy the monopoly, since he can reproduce the information at little or no cost. Thus the only effective monopoly would be the *use* of the information by the original possessor. This, however, will not only be socially inefficient, but also may not be of much use to the owner of the information either, since he may not be able to exploit it as effectively as others.

With suitable legal measures, information may become an appropriable commodity. Then the monopoly power can indeed be exerted.²¹

¹⁸ SUSANNE SCOTCHMER, "Standing on the Shoulders of Giants: Cumulative Research and Patent Law," 5 *Journal of Economic Perspectives* No. 1, Winter 1991, at 29-41.

¹⁹ Cf. FRITZ MACHLUP, "Patents, *International Encyclopaedia of the Social Sciences*" 461-472 (The Macmillan Company & The Free Press, 1968)

²⁰ 35 U.S.C. § 154 (1988).

²¹ ARROW, *supra* note 2, at 615.

It is not quite clear what Arrow actually means by “the *use* of the information.” In any case, patents are not such “suitable legal measures” by which an “effective monopoly” could be established in the sense that the “*use* of the information by the original possessor” would also include the right “to sell information on the open market.” Such a monopoly is a priori impossible, given the inherent public-good effect of patents. In fact, patents destroy any chance of such a market – which however must not be confused with the market for related innovative products. Consequently, Arrow’s “suitable legal measures” cannot be interpreted to mean the established patent system, at least not without some qualifications, although Arrow does briefly mention patents: “The property rights may be in the information itself, through patents and similar devices....”²² Here, he makes no distinction between the private-good effect and public-good effect of patents, and the claim that patents and similar devices may create “property rights in the information itself” is at best only true as far as the embodiment of information into related products is concerned.

Another example is the more recent reasoning by Stiglitz:

Initial knowledge is a key input into the production of further knowledge, and thus the design of the patent system can dramatically affect the overall pace of innovation. An excessively broad patent system (for example, with long-lived patents of broad scope) can raise the price of one of the most vital inputs into the innovative process and thus reduce the follow-on innovations, even as it may provide returns to those making the original innovation. As a result the overall pace of technical progress may be slowed.²³

Stiglitz is raising a highly relevant issue, yet his explanation of the possible impact of patents on technical progress is misleading and needs clarification, because the public-good effect of the patent system works in exactly the opposite way. Again and again, the price of patent information used as “initial knowledge as a key input into the production of further knowledge” is zero, irrespective of the scope and duration of patent protection. As already noted, this fact as such can obviously only accelerate overall technical progress. Whether the benefits so generated are subsequently impaired by excessive protection of “long-lived patents of broad scope,” *i.e.*, by the private-good effect of patents, is certainly a key issue, yet the correct explanation is different from what Stiglitz asserts.

Another important source of the inadequacy of the mainstream doctrine has been briefly mentioned already, namely the misconception about an overly broad scope of protection. The view that patents protect both ideas and discoveries in contrast to copyright, which only protects the expression of ideas, is widely held.²⁴ If this were true, then without any doubt patents would indeed be very strong monopolies. Yet such a view is simply not correct.

²² ARROW, *supra* note 2, at 617.

²³ JOSEPH E. STIGLITZ, “Knowledge as a Public Global Good,” in: INGE KAUL, ISABELLE GRUNBERG & MARC A. STERN (eds.), “Global Public Goods” 312 (Oxford University Press, Oxford 1999) (published for the United Nations Development Programme).

²⁴ *Cf.* JOHN PALMER, “Editor’s Introduction,” in: “Research in Law and Economics,” *supra* note 1, at 133-4; CHRISTOPHER MAY, “A Global Political Economy of Intellectual Property Rights” 7 (Routledge, London 2000).

First and foremost, there exists an analogy in patent law to the famous idea/expression dichotomy in copyright: scientific discoveries and/or laws of nature in patent law may be compared to “ideas” in copyright law, because as such they are a priori not patentable; only their “expression” in the form of one or several inventions possessing the ability to be reduced to practice is eligible for patent protection.²⁵ Despite being almost unnoticed, this analogy is highly relevant, because the idea/expression dichotomy is in economic analysis correctly interpreted as the pro-competitive (or non-monopoly) feature of copyright.²⁶ The observed analogy then naturally implies that patents, much like copyright, cannot be *a priori* considered as a monopoly and most of the already quoted evidence²⁷ solidly supports such a conclusion.

In this respect, it is also a fact that there is formidable semantic confusion about the meaning of the word “idea” itself. Virtually all economists agree that innovative ideas are patentable, although only some of them attach relevant attributes. Maskus, for example, correctly speaks about “ideas with industrial applicability,”²⁸ which manifestly makes an enormous distinction to widespread claims that ideas are generally patentable.²⁹ On the other hand, legal scholars unanimously claim exactly the opposite. The following remark by Cornish is a representative case in point: “For instance, in an internal combustion engine, the *idea* of putting a cushion of air in the cylinder between the fuel and the piston in order to cushion the explosive effect of ignition was said not in itself to be patentable; but a machine devised to do so was.”³⁰ Much the same reasoning was expressed in the U.S.A as far back as 1874 in the *Rubber-Tip Pencil Co. v. Howard* case, where the court stated that “An idea itself is not patentable, but a new device by which it may be made practically useful is.”³¹ One must thus be careful what the term “idea” actually means when a claim is made about patentability or non-patentability.

III. The Competitive Incentive To Invent

The above discussion suggests what needs to be undertaken in order to find a more appropriate explanation of how the patent system works in a modern economy. The assumption that incentive to invent equals monopoly profits must be replaced with an alternative that would be sufficiently appropriate to reflect the prevailing knowledge-based competition. This would then be the basis for the main objective of developing a different, presumably more adequate basic economic rationale of patents, which must be of course also properly interpreted in respect of what they protect and what they not protect.

²⁵ Cf. DAVID A. BURGE, “Patent and Trademark Tactics and Practice” 36 (John Wiley and Sons, Inc., New York etc. 1999).

²⁶ WILLIAM LANDES & RICHARD POSNER, “An Economic Analysis of Copyright Law,” XVIII Journal of Legal Studies, June 1989, at 361.

²⁷ *Supra* note 8.

²⁸ MASKUS, *supra* note 5, at 39.

²⁹ PALMER, *supra* note 24.

³⁰ WILLIAM R. CORNISH, “Intellectual Property: Patents, Copyright, Trade Marks and Allied Rights” 207 (Sweet & Maxwell, London 1999) (emphasis added).

³¹ Cf. DONALD S. CHISUM & MICHAEL A. JACOBS, “Understanding Intellectual Property Law” 2-28 *et seq.* (Matthew Bender, New York 1992).

The key issues are thus how to replace the current concept of the incentive to invent with a more suitable alternative and how to interpret the economic impact of patents thereunder. As long as Schumpeter's postulate about monopoly profits as the only possible reward for inventive activity is assumed to be true, then even the most correct interpretation of patent law is of no help in moving away from the patent-is-monopoly view, since it is an integral part of the assumption itself.

What could then be an alternative incentive to invent? Nowadays, it is a notorious fact that, without innovation, it is impossible to stay in business in any industry in which innovation is, to use Baumol's terminology, a routinized activity. And there is hardly any industry in which routinized innovation is not the rule of the game. Now suppose that a firm within a given industry cannot or does not want to innovate, for whatever reason. Clearly, it can then compete solely on the basis of obsolete technology, a very weak and short-lived alternative indeed. After a time, any non-innovating firm will be faced with "considerable erosion of its markets"³² and will eventually die, as Freeman and Soete state.³³

Therefore, there is no question of whether or not to innovate, but rather in which direction and how fast. All firms must innovate if they wish to survive, and this fact can easily be identified as the basis for an alternative concept of incentive to invent. The following definition – conveniently termed here as the "*competitive incentive to invent*," in contrast to Schumpeter's "*monopolistic incentive to invent*" – is therefore at hand: *The competitive incentive to invent aims to prevent negative profits (losses), which would inevitably occur in the absence of continuous innovation.*

This proposed definition requires two brief remarks. Firstly, it would actually be more appropriate to speak of (inescapable) "innovation pressure" rather than an "incentive." Despite this, the established notion of "incentive to invent" (more correctly, "incentive to invent and innovate") has been retained for terminological convenience. Secondly, the competitive incentive to invent does not necessarily stand as a mutually exclusive alternative to the classic monopolistic incentive. An early innovator may strive to achieve a kind of competitive advantage in terms of a monopoly. However, this is precisely what others want to prevent with their own innovative efforts, but what may occasionally fail for whatever reason. If such a failure occurs, then a monopoly may eventually be the outcome, yet this would clearly be a rare exception and certainly not the rule. The competitive outcome remains the rule, and thus the proposed competitive incentive to invent nowadays generally prevails, with the consequence that "[T]he profit-maximizing firm will adopt the quantity of R&D expenditure at which expected marginal profit is zero."³⁴

IV. Patents as the Means Against Free Riding

We may now examine the key issue, *i.e.*, a possible principal economic function of patents within the framework of the competitive incentive to invent. Intuitively, it could be predicted that patents might become irrelevant in an innovation-based competitive environment. If all firms must continuously innovate in order to survive, why should they seek patent protection at all?

³² BAUMOL, *supra* note 15, at 45.

³³ CHRIS FREEMAN & LUC SOETE, "The Economics of Industrial Innovation" 266 (Pinter, London 1997).

³⁴ BAUMOL, *supra* note 15, at 45.

Suppose that innovating firms are established in a given industry. For them, R&D outlays must be considered to be a part of the overall costs. Yet there may also be other firms, especially potential entrants to the industry, that have not, or not yet, engaged in their own R&D projects. And here patents enter the scene. In the absence of patent protection, any innovation can be copied and freely commercially exploited by another, non-innovating firm, which therefore does not need to spend on its own R&D. Consequently, minimum average and marginal costs are then lower for the non-innovating firm, which thus enjoys a cost advantage over *all* innovating firms, *ceteris paribus*.

In other words, in a world in which there is no patent protection, innovating firms are faced with the phenomenon of free riding. If free riding occurs, then innovating firms are literally punished. In the worst case, they may eventually be driven from the market because their own expensively developed inventions can be copied by free riders at much less or even no cost. In terms of entry barriers to the market, this would mean that innovating incumbents, not free riders, would face a barrier, *i.e.*, costs spent on R&D; a paradoxical situation indeed.

This would suggest that the basic economic function of patents is different from that postulated in the mainstream doctrine. If patents prevent free riding, it may be assumed that *only* free riders are excluded from competition, but *none* of the other firms that wish to introduce their own innovations. Assuming also that all innovating firms eventually face equal or at least very similar cost curves in both production and R&D activities, then a sustainable competitive market structure would be maintained. If many innovating firms exist, competitive price-taking behavior would prevail, with the R&D investments of each and every firm still being fully recovered. Such a market structure could conveniently be termed as "*innovation-based competition*," the notion already used but not formally explained.

Accepting for the moment the above assumptions as valid, then the principal economic function of the patent system can be summarized in one single sentence: *The principal economic function of patents is to establish innovation-based competition solely by limiting any cost advantage on the basis of free riding.*

To some extent, the proposed economic function of patents is not completely original. For example, Reichman takes a similar standpoint, as already reflected in the title of his study on the role of the TRIPS Agreement in global competition.³⁵ However, the methodology and arguments chosen here are significantly different and provide the important benefit of new insights and implications. Furthermore, a formal microeconomic model of this "new" function of patents can be constructed, as shown below.

Before continuing the analysis, however, we must first ascertain whether the main assumptions behind the suggested principal function of patents are sufficiently reasonable from both theoretical and empirical standpoints. Therefore, the impact of *the scope of protection* as determined by law must be examined. As already noted

³⁵ JEROME H. REICHMAN, "From Free Riders to Fair Followers: Global Competition Under the TRIPS Agreement," 29 New York University Journal of International Law and Politics Nos.1-2, Fall 1996-Winter 1997, at 11-93.

above, it is necessary to check that the pro-competitive nature of the public-good effect of patents is not hampered, wholly or partly, by their private-good effect. Prevention of free riding as such clearly does not *a priori* imply that patents cannot establish a monopoly. If patent protection is excessively broad, then not only free riders but also all other R&D-spending competitors may be faced with entry barriers in terms of the relatively higher costs of R&D. This is the essence of the issue raised by Stiglitz, but set in an appropriate context here.

Generally speaking, the scope of patent protection is basically limited by law in two different ways.

First, there is an absolute exclusion of scientific discoveries and laws of nature as such from patent protection. As discussed above, this requirement, which resembles the non-protectability of ideas in copyright, is unambiguously met everywhere in the world, being one of the most fundamental principles in patent law. For example, discoveries, scientific theories and mathematical methods as such are explicitly excluded from patentability by Art. 52(2) of the European Patent Convention, although this is not the case with the ubiquitous WTO TRIPS Agreement, reflecting only a different logic but not a change in substance; scientific discoveries or laws of nature are by definition not inventions and there is thus actually no reason to exclude them explicitly.

This crucial limitation of the scope of patentable subject matter *a priori* leaves the way open to many different competing inventions that can be developed on the basis of freely available scientific knowledge. This fact, solidly confirmed by empirical evidence,³⁶ is additionally reinforced by the public-good effect of patents, *i.e.*, the information contained in valid patents is freely available in the same way as any other non-patented information. Therefore, the stock of all the freely available knowledge as an input into the creation of new knowledge is significantly expanded. Recall the well-known fact that exactly this information is extremely valuable, because each and every patented invention must be disclosed up to the point that would “enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same.”³⁷ Moreover, patents and/or patent applications are nowadays published just 18 months after the first filing date in all major countries. Therefore, information about the most recent inventions is available and thus the risk of infringing competitors’ patents can be minimized in a timely manner.

The second limitation is of a slightly different nature and concerns the issue of how broad the scope of protection for each patented invention may be. Ideally, the scope of protection ought to be limited to the extent that does not force competitors to invest, in relative terms, more into the R&D of their own alternative inventions, *ceteris paribus*. If this is achieved, then only free riding is prevented, but not innovation-based competition. On the other hand, if the scope of granted patents is too broad, then entry barriers are erected regardless of their public-good effect. By the same

³⁶ Cf. SCHERER & ROSS, *supra* note 4, at 624-5. One notable exception in this respect are biotechnological inventions, in which it is difficult to draw a clear-cut borderline between invention as such and the underlying scientific discovery or law of nature. Yet this issue can be resolved by law.

³⁷ 35 U.S.C. § 112; see also, *e.g.*, Art. 83 of the European Patent Convention and/or Art. 5 of the WIPO Patent Cooperation Treaty.

logic, if the scope of protection is too narrow, some degree of free riding is likely to occur.

In contrast to the uniform exclusion of scientific discoveries from patent protection, the legal situation in this respect is somewhat more complicated. In principle, the actual scope of protection is determined by patent claims everywhere in the world; however, there are considerable differences among various states as far as legal rules and judicial practice of interpretation are concerned.³⁸ This fact obviously calls for global harmonization of the norms concerning interpretation of patent claims; nonetheless, the main point is that the whole set of legal rules and practice concerning the scope of patent claims and its interpretation is crucial to determining the impact of the patent system on market structure. This is a conclusion of immense importance for policymakers, simply because it is practically feasible, which can hardly be said for some other theoretical policy recommendations.

Despite this indeed laconic, overly simplified and incomplete presentation of the basic principles, it is not difficult to see that the scope of patent protection is *itself* limited by the law to the extent that the primary benefit of the public-good effect of patents is not adversely affected.

Therefore, we may conclude that the crucial assumption, *i.e.*, to consider the basic economic function of the patent system as a means solely to prevent free riding without impairing innovation-based competition, has been found to be compatible with most fundamental principles of patent law.

It must be added that the assumption, framed so far in an abstract theoretical context, is strongly supported in reality, too. With few exceptions, a huge number of patented and unpatented alternative competing solutions for essentially the same purposes can be observed in most industries. Empirical studies by Levin et al. and by Mansfield³⁹ show that patents in most cases allow not only innovation-based competition, but even some degree of free riding. Note that at the time when these studies were carried out, the early disclosure of patents, an important pro-competitive feature, was not yet practiced in the U.S.A. Patents have therefore been found to be of even less than “optimal” scope; optimal in the sense that the scope of protection would require competitors to keep their R&D investments exactly at the same level as that of other patentees, *ceteris paribus*.

V. The Number of Competing Patents and Term of Protection

The approach as developed above offers some interesting theoretical insights in respect of the possible number of patents and their duration.

As far the number is concerned, the suggested economic function of patents implies that *each and every firm must protect all of its inventions*, so that none of them would be made available cost-free, or at least less costly, to free riders. In theory, there must eventually be precisely as many patents as the number of competing products or processes that have been independently developed and put on the market. This

³⁸ See, e.g., TOSHIKO TAKENAKA, “Interpreting Patent Claims: The United States, Germany and Japan” (VCH Publishers Inc., New York 1995).

³⁹ *Supra* note 8.

outcome is the extreme opposite of the outcome predicted under monopolistic incentive to invent and under single-prize, winner-takes-all patent races. Although both outcomes are of a purely theoretical nature, it is immediately clear that the outcome implied under the competitive incentive to invent is much closer to observed reality than the currently prevailing economic doctrine; patent statistics are more than convincing evidence in this respect.

With regard to the term of protection, this analysis again implies yet another entirely different outcome than that in traditional theory. Since patents are not a monopoly, they obviously cannot be taken to be a trade-off between static and dynamic efficiency. This in turn implies that Nordhaus's pioneering work on the optimal life of patents⁴⁰ seems to be of no direct relevance here, because time as a variable simply does not appear. If so, then the impressive stock of subsequent research on this basis must also be put into question, including the contributions of Gilbert and Shapiro⁴¹ and of Klemperer,⁴² which are worth mentioning, because they arrive at exactly the opposite results to each other. By the same token, the policy recommendation "to tailor the life of each patent to the economic characteristics of its underlying invention"⁴³ is also losing ground, apart from being immensely complex and thus highly impractical.

Yet an important caveat must be made here. The disappearance of time as an economic variable does not warrant the far-reaching conclusion that patent protection could be extended to last indefinitely, for the following two main reasons.

Firstly, firms in any given industry differ significantly to each other in size, technology, managerial strategy, etc., all over the world; the ideal textbook atomistic market structure in which all firms are identical in each and every respect does not exist. Thus, in reality many firms with their own innovations are not able to compete *on an equal basis* with more advanced and/or bigger competitors. However, the relative lack of innovative efforts may be compensated somehow, say, by lower wages or in some other ways, *but only as long as the innovative gap is not too large*. This means that patent protection must not be extended over an excessively long period. The adequately limited duration of patents is thus an important remedy, which gives many firms the chance to stay in business despite their R&D disadvantage relative to innovation leaders. A relatively higher number of firms may consequently be present on the market, and all the beneficial effects – marginal pricing, etc. - of such a competitive market structure should therefore be assured. A typical and illustrative case in point is the pharmaceutical industry with relatively few large innovating firms and many generic manufacturers of smaller size, who simply cannot afford the exorbitant R&D investments required for successful development of new drugs. Consequently, generic firms must often wait for innovators' patents to expire and thus enter the market at a later point of time.

⁴⁰ WILLIAM D. NORDHAUS, "Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change," Ch. 5 (The MIT Press, Cambridge, Massachusetts 1969).

⁴¹ RICHARD J. GILBERT & CARL SHAPIRO, "Optimal Patent Length and Breadth," 21 Rand Journal of Economics No. 1, at 106-112 (1990).

⁴² PAUL D. KLEMPERER, "How Broad Should the Scope of Protection Be?," 21 Rand Journal of Economics No. 1, at 113-130 (1990).

⁴³ FREDERICK M. SCHERER, "Innovation and Growth" 139 (The MIT Press, Cambridge, Massachusetts 1984).

Therefore, the fundamental principle of the patent system that all patented inventions must eventually fall into the public domain is not only relevant, it is of utmost economic importance. The term of protection must be very carefully determined, so that the innovative gap does not grow beyond the point where the survival of too many smaller firms would no longer be feasible. This criterion needs to be considered whenever claims are made that the term of protection is too short for a given (innovative) industry, because such claims usually support only one side of the issue.

Secondly, it was mentioned above that the competitive incentive to invent does not exclude the possibility that, exceptionally, a monopoly may be established. The history of patents indicates that this may happen especially when major technical breakthroughs occur, so that such monopolies, if established, may indeed have quite an economic impact. In the past, one single patent covered a whole car, airplane, locomotive or steamship; nowadays, a similar pattern can still be observed. For example, as far as the emergence of biotechnology is concerned, Grubb says that “[B]oth the USPTO and the EPO have at times granted biotech patents with excessive width.”⁴⁴ Under such circumstances, it again makes sense to be cautious and to limit the duration of patent protection.

It is worth noting, however, that the above reasons for limiting the term of patent protection are conceptually completely different from those postulated in mainstream theory. Whether this fact calls for a reconsideration of the optimal term of patent protection is an issue that is not pursued here.

VI. A Diagrammatic Exposition

The above deliberations can be translated into a relatively simple economic model, albeit with some further simplifications, which, however, do not alter the basic conclusions. The model deals with cost-reducing innovations only, which makes it suitable to be compared with Arrow’s model in the geometric version of Demsetz.

Suppose that there is competitive supply initially. All competing firms are identical, producing a homogeneous product, with identical cost curves. The total cost curve TC is given as the sum of fixed costs F and increasing variable costs C(q), thus $TC = F + C(q)$, $dC/dq > 0$. The average total cost curve ATC is then obtained by dividing TC by the quantity q, $ATC = F/q + c(q)$, or $ATC = AFC + AVC$, where $AFC = F/q$ and $AVC = c(q) = C(q)/q$. The ATC curve then takes a familiar U-shaped form, as depicted for a representative firm in Figure 1. Given the competitive supply, firms are price takers and sell at a single competitive price P, that is, where positively sloped marginal cost MC equals the minimum average total cost ATC. After innovation takes place, the respective costs and quantity are denoted by the subscript i.

⁴⁴ PHILIP W. GRUBB, “Patents for Chemicals, Pharmaceuticals and Biotechnology” 237 (Clarendon Press, Oxford 1999).

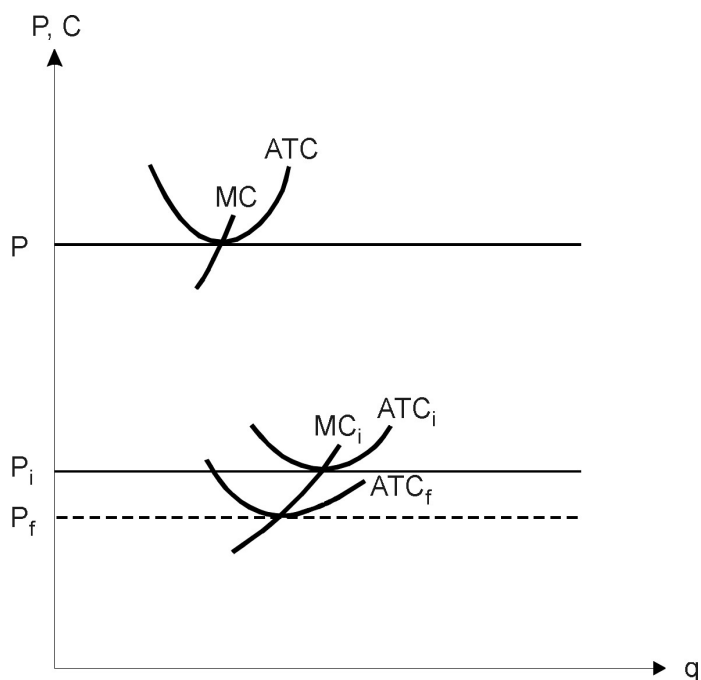


Fig. 1

All firms spend the same amount, denoted RD, on research and development. The incentive to undertake research and development arises from the need not to lag behind the others, as competition using obsolete technology, *i.e.*, higher costs, is not feasible in the long run.

The amount RD may either be taken as a fixed cost or as some fraction of unit costs. The first option is convenient for the U-shaped costs as shown in Figure 1. The second option, shown in Figure 2, is practical in the case of constant unit costs, the case analyzed by Arrow.⁴⁵ In either case, every firm spending on R&D succeeds in developing its own cost-reducing invention, which makes production technology cheaper but leaves the manufactured product unchanged. The achieved reduction of costs is the same for all firms, although the inventions themselves differ from a technical point of view. This assumption is somehow compatible with the analysis by Needham.⁴⁶ Note that this does not also mean a wasteful, Pareto inefficient duplication of R&D efforts, as the outcome of each R&D effort is a different invention, and thus the stock of knowledge available for (free) use in the further production of knowledge is significantly enhanced.

It is noteworthy that the above – and for this model crucial – assumptions are in full compliance with standard legal provisions of patent law.

It is further assumed that each and every invention meets the usual patentability criteria, and patent protection is granted to each and every firm for its invention, whereby the use of such an invention does not constitute an infringement with regard to other inventions developed and patented by competitors. Finally, it is assumed that

⁴⁵ ARROW, *supra* note 2.

⁴⁶ DOUGLAS NEEDHAM, "Market Structure and Firms' R&D Behaviour," XXIII The Journal of Industrial Economics No. 4, June 1975, at 241-255.

all firms innovate simultaneously. This is the logical consequence of the assumption of equal RD spending, and also because all the information about competitors' patented inventions is nowadays publicly available at a very early stage, in most cases even before the beginning of commercial exploitation.

The obvious effect of all inventions under the above assumptions is that the new U-shaped post-invention ATC_i curve (Figure 1) is the same for all firms. In the relevant range of the quantity produced, the new curve exhibits a lower minimum than the ATC curve, $ATC_i < ATC$, as shown in Figures 1 and 2. However, the amount of RD must be included in the post-invention average fixed costs AFC_i (not drawn). Since other fixed costs are assumed to remain unchanged before and after the innovation, post-invention average fixed costs AFC_i are greater than initial fixed costs AFC . This in turn implies that the innovation must result in such a reduction of variable unit costs $AVC=c(q)$, which more than offsets the increase in fixed costs. Therefore, $AFC+RD/q = AFC_i > AFC$, but $c_i(q) \ll c(q)$, so that eventually $ATC_i < ATC$ in the relevant neighborhood of their minima. This then also implies lower marginal costs after innovation: $MC_i=dC_i/dq < MC=dC/dq$.

After all firms innovate, a cost symmetry is established in exactly the same way as was the case before the innovation. Consequently, the competitive equilibrium is fully re-established, albeit at a lower price and cost level. All firms have to continue to behave as price takers, by setting their prices to the new, lower competitive price P_i .

This outcome holds true as long as *all* innovating firms obtain patents for *all* their inventions. To illustrate this, suppose that an innovating firm (or several of them) fails to apply for a patent. Then any third firm, possibly a potential entrant to the market, is legally allowed to free ride. If this happens, then virtually no legal remedy remains available to the firm that had failed to file for patent protection on time, because novelty has been lost.

The copying costs CC of a free rider – or several of them - are in most cases smaller when compared to the RD outlays of an innovator. At the extreme, copying costs CC are zero.⁴⁷ Consequently, a free rider faces a different average total cost curve than a firm that innovates on the basis of RD spending. This curve, denoted ATC_f in Figures 1 and 2, obviously lies below the ATC_i curve. Clearly, the difference between the two curves is then the cost advantage enjoyed by free riders, *ceteris paribus*. They can always undercut the lowest competitive price P_i of the RD-spending firms. Note that a single free rider cannot normally become a monopolist, because there is always a possibility for many free riders to enter the relevant market. This fact may actually drive the price close to P_f , the price that innovating firms cannot match without losses.

In other words, in the absence of even a single patent for any innovation, the market is eventually overtaken by free riders, whereas innovators remain empty-handed. It may be safely claimed that, in reality, innovating firms are aware of the need to obtain patent protection for this purpose, powerful evidence being the high numbers of patents filed by many firms competing with each other. On the other hand, patents

⁴⁷ Here we speak strictly of "copying." The frequently used expression "imitation" is ambiguous, as it may mean parallel R&D activity, which leads either to identical inventions or to technically different inventions that serve the same economic purpose.

thus only protect innovating firms from competition based on free riding, but do not reward innovators with a monopoly because of legal limitations on patentable subject matter. Innovation-based competition is therefore open and feasible for all firms that are willing and capable to spend money on their own research and development.

The same logic is almost directly applicable to the constant unit case, but with a slight change. As already indicated, it is more convenient in this case to assume that firms are spending a certain and constant fraction of the unit costs, denoted RDC, for the purposes of innovation. It is then easy to see that a free rider would enjoy a cost advantage equal to that particular cost fraction RDC, if the copying costs CC for an unpatented invention were zero. If copying costs CC are positive, then the cost advantage of a free rider is of course smaller. More simply, $P_f = P_i - RDC$, if $CC = 0$. If $CC > 0$, then $P_f = P_i - RDC + CC$, of course as long as $RDC > CC$.

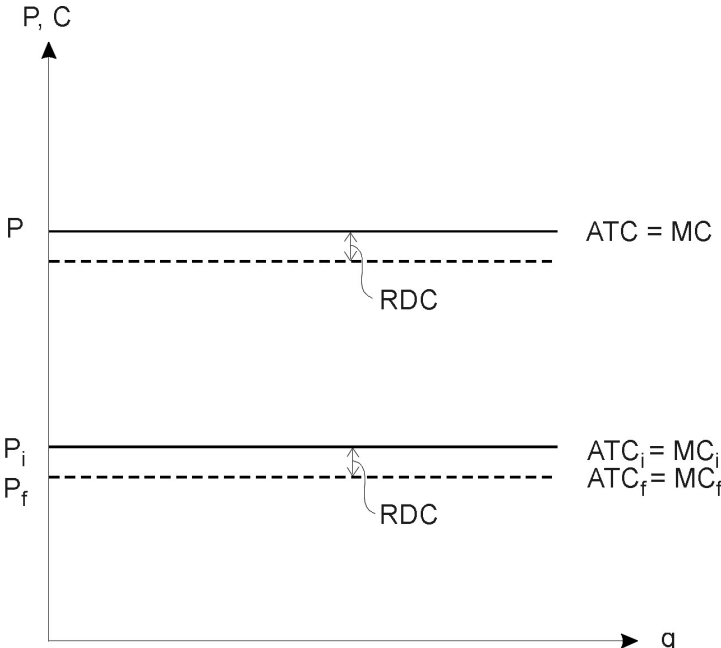


Fig. 2

Assuming product homogeneity and cost symmetry, the outcome is Pareto efficient. In cases involving new products, however, the homogeneity assumption cannot hold. Then Chamberlin’s classic tangency solution of monopolistic competition would be the outcome under the main features of the model. Such an outcome is nonetheless still Pareto superior when compared to the pure patent monopoly.

Finally, it may be useful to say a few words about the issue of the optimal scope (breadth) of patent protection within the model. Note at the outset that, in contrast to the prevailing theory that inevitably sees patents as a trade-off between static and dynamic efficiency, time as a variable does not appear at all in our approach and thus plays no role in determining the optimal scope of patent protection.

As before, let RD denote the amount spent on the development of a new invention, and CC the costs of copying this invention if left unprotected. In an optimal case,

patent protection should create a situation in which any free rider would face additional costs that would exactly fill the gap between copying costs CC and RD outlays spent by innovating firms. If so, then there is no longer an incentive to copy, as the cost advantage on the basis of free riding disappears. Moreover, copying a patented invention would become a risky affair, because it would inevitably result in patent infringement. The only alternative left then is to innovate (or not to enter the market, but we leave this alternative aside).

As stated, it is assumed that a competitive innovation can be made with a level of RD spending that is not higher than the spending of other innovators. In terms of a simple equation, the sum of (optimal) costs incurred by patent protection, denoted as SP , and copying costs CC must equal the RD , *i.e.*, the amount spent on innovation by each firm. This is the same as saying that SP is the difference between RD and CC costs:

$$SP = RD - CC$$

or

$$SP = (RD/CC - 1).CC$$

The term in brackets can be seen as a kind of multiplier by which the copying costs must be multiplied in order to prevent free riding. If the equation holds, then the optimal scope of patent protection is achieved.

The value of the term in brackets reflects the technology in question. In most cases, *i.e.*, when copying is less costly than initial research and development, the value is of course greater than one; it may even reach values over 100. This may frequently occur in the case of new medicines, and explains why patents loom so large in the pharmaceutical industry. Clearly, the greater the value in brackets, the greater the (relative) importance of patent protection.

However, the value of RD/CC may take, in principle, a value close to or even less than one, implying that copying costs CC are equal or even greater than development costs RD , $RD/CC \leq 1$. Thus, $SP \leq 0$ (negative values are, of course, meaningless). It is certainly a rare case where patent protection, or any other similar form of legal protection, does not matter, because free riding is a priori ruled out for technological reasons. An example seems to be the semiconductor industry. Technically it is possible to discover the topography of a chip for the purpose of copying it, but the process is usually more costly than the comparatively cheaper development of a new chip with the same functional features. This fact may explain why the number of protected topographies (which enjoy a *sui generis*, quasi-patent protection) is low relative to the huge output of integrated circuits worldwide.⁴⁸

In sum, we can then establish the following simple rules for the optimal scope of patent protection:

$$SP_{opt} \cong (RD/CC - 1).CC, \text{ for all values of } RD/CC \geq 1$$

⁴⁸ Cf. GUNNAR W.G. KARNELL, "Protection of Layout Designs (Topographies) of Integrated Circuits – R.I.P.?", 6 IIC 648 (2001).

and

$$SP_{\text{opt}} \cong 0, \text{ for all values of } RD/CC \leq 1.$$

These rules are depicted in Figure 3.

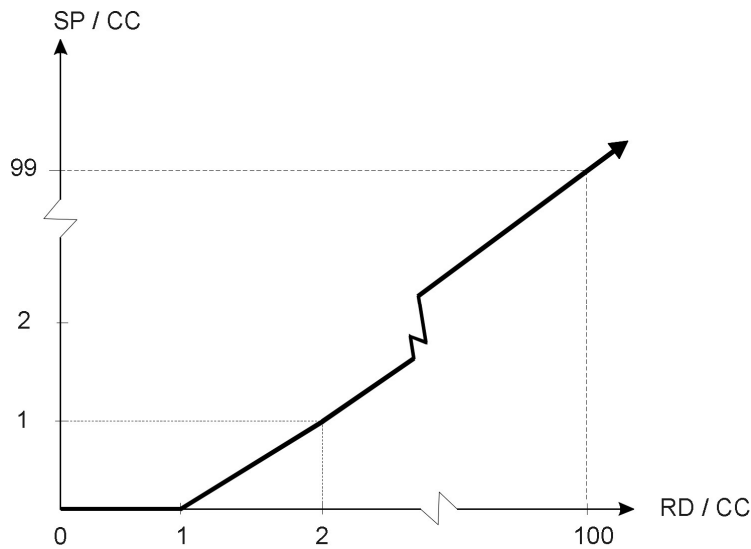


Fig. 3

The above exposition of the competitive framework of patents and their optimal scope is an extremely simple model, the value of which rests primarily on the originality of the underlying assumptions and not on its sophistication. However, it is clear that the story is of course much more complex. Two examples may illustrate the point. First, firms do not generally invest the same amounts of RD, although it is then difficult to say whether this difference is a consequence of non-optimal scope of patent protection (non-optimal in either way) or of deliberate efforts to invent “more” and/or faster, with the aim of achieving some initial monopoly. Secondly, one could postulate that if a certain number of inventions were already made, then a relatively larger amount of RD would be necessary to catch up. This fact could lead to an excessive, socially suboptimal RD investment, as firms would tend to race for relatively less expensive initial patents, which would potentially establish an entry barrier. All these, and of course many other, issues have already been extensively analyzed, although in the conceptually different framework of traditional doctrine. Nonetheless, they may be susceptible to further analysis of the competitive paradigm of patents as exposed here.

VII. Concluding Remarks

The main conclusion of this paper is that patents, if properly interpreted, do not create monopolies for innovators. Patents, being simultaneously private and public goods, only prevent competition based on free riding, but at the same time they allow, even support, the spread of innovation-based competition, whose main driving force is what was named here as the competitive incentive to invent.

Despite its simplicity, the suggested basic logic reflects the current reality much more accurately than the traditional view based on the outdated assumption of Schumpeter's incentive to invent.

In the classic study "An Economic Review of the Patent System," Fritz Machlup⁴⁹ came to, in his own words, a "disappointingly inconclusive conclusion" on whether a society would be better off with or without the patent system. In fact, Machlup was not able to find a clear answer to the question that involved two extremes. The first extreme could be labeled as the "full" private-good effect of patents – which is incorrectly often assumed actually to exist – without any public-good effect, meaning that the free use of knowledge disclosed in granted patents would not be allowed. Such a system would without any doubt make patents an effective monopoly. The other extreme is of course the "full" public-good effect, *i.e.*, a situation achieved by complete abolition of the patent system, which would most probably lead to a slowdown of technological progress. But it should be remembered that Machlup also follows the view that monopoly profits are the sole source of incentive to invent.

However, under the concept of competitive incentive to invent the possible conclusion whether to have a patent system or abolish it is no longer so disappointingly inconclusive. A patent system with the proper mix of private-good and public-good effects seems to be an obviously attractive and viable alternative to the two extremes: there is no absolute monopoly with watertight entry barriers and competition is open to all those firms that are willing (and able) to innovate, but not to free riders. Thus the *nature* of competition is affected by such a patent system, but competition itself is in no way eliminated.

Such a conclusion is obviously distant from the widespread view that the patent system is an unpleasant economic compromise, by which a temporary monopoly has to be established simply because there is no better alternative to promote technical progress. As a corollary, this analysis offers simple policy guidance: the basic features of the patent system need to be preserved as they are. In other words, lawmakers must be careful not just blindly to follow demands by industrial lobbies that would like to see patent protection extended in one way or another to scientific discoveries, and/or to impose limitations on the free use of knowledge disclosed in valid patents. In addition, they must take care that the term of protection would not be prolonged beyond a reasonable period. Last but not least, care must be taken that the scope of patent claims and their interpretation does not become excessively broad. Under such sound circumstances, only free riders but not fair followers, to paraphrase Reichman, will continue to face barriers to taking part in innovation-based competition.

⁴⁹ FRITZ MACHLUP, "An Economic Review of the Patent System," Study No. 15, Subcomm. Patents, Trademarks and Copyrights of the U.S. Senate Judiciary Comm. (U.S. Government Printing Office, Washington, D.C. 1958).