Chapter 10

ARE WE ON OUR WAY IN THE NEW ECONOMY WITH OPTIMAL INVENTIVE STEPS?[1]

Ove Granstrand
Chalmers University of Technology, Department of Industrial Management and Economics, Göteborg, Sweden

Chapter contents:

10.1	Purpose and Outline	224
10.2	Are We on Our Way into a New Economy?	225
10.3	What Role has the Intellectual Property System Played?	226
10.4	What are the Challenges to the IP System?	228
10.5	What Remedies are There to the IP Assembly Problem?	235
10.6	Could the Inventive Step Requirement be Raised?	237
10.7	Optimal Inventive Step in the Simplest Case	240
10.8	"Evergreening" by Incumbents	247
10.9	Alternatives to Patents	252
10.10	Summary and Conclusions	255
10.11	Literature References	256

Abstract: As the economy is shifting weight from physical to intellectual capital dominance, accompanied by the emergence of the pro-IP era, it is natural to try to reassess the traditional challenges, trade-offs and operating standards of the IP system. There is no change in the fundamental nature of information and technology with its associated legal and economic difficulties in arranging for property rights, investment and trade, but the proportions of various difficulties increase with the increasing role of information in the economy. This

[1] I want to thank John Barton, Lars Björklund, Tom Ewing, Jakub Gubanski, William Kingston, Joseph Straus and Hanns Ullrich for helpful discussions and comments on this chapter. The competent research assistance of Ola Küller is greatly appreciated. The help from patent offices and examiners interviewed in Japan, US, Denmark and Sweden is also gratefully acknowledged. This chapter is a partial result from a study of patentability criteria, especially the inventive step requirement. The financial support from The Romanus Fund and the Research Fund of the European Patent Organisation is also gratefully acknowledged.

chapter addresses the causes and consequences of the anti-commons problem with fragmentation of interdependent IPRs in complex technologies and the associated problems for IP assembly and technology trade. In particular the chapter analyses patentability criteria and the possibilities to use the inventive step or non-obviousness standard as one out of several policy remedies. Raising the minimum inventive step standard will reduce transaction costs, invent-around costs and hold-up problems in commercializing innovations, but will also reduce incentives to innovate under certain conditions. While an optimal inventive step standard exists in principle, and uniquely so in certain cases, its empirical determination is obviously a non-obvious matter. However, a low inventive step standard may be instrumental for incumbent "evergreening", i.e. prolonging effective patent protection, and bridging technology transitions. The patent system could finally be studied as a decentralized taxation scheme in a governance context, taking total governance costs into account.

Keywords: Patents, IPR, anti-commons, IP assembly, inventive step, transaction cost, technology markets

10.1 Purpose and Outline

If the economy is gradually shifting weight from physical to intellectual capital dominance, it is natural to try to reassess the traditional challenges, trade-offs and operating standards of the IP system. The purpose of this chapter is to ask, first, how the IP system is challenged by a new type of economy, and how transaction costs (or more generally governance costs) could be taken into account when analysing the economics of IPRs. In particular the chapter elaborates upon the causes and consequences of difficulties in assembling the necessary IPRs in complex technologies with many agents, possibly resulting in excessive and/or deterring transaction costs. A subsequent purpose of the chapter is to discuss how this problem could possibly be mitigated by raising the inventive step requirement in granting patents. A first attempt is thereby made to address the issue of optimality of an inventive step requirement in a static framework and then in a dynamic situation involving sequential innovation and a technology transition. In that connection the chapter cannot answer the question raised in its title, but just elaborate upon it a bit. Finally, the chapter briefly views the patent system in a governance context, in which total governance costs with both transaction costs and management costs have to be taken into account.

10.2 Are We on Our Way into a New Economy?

There are several sound reasons to be skceptical about much ado over the so-called new economy. Is the concept of a new economy just another management fad, boosted by a new generation of impatient entrepreneurs, opportunistic consultants, exaggerating media, pro-active stockbrokers and the like? Is it a product of the propensity to announce deep-running new changes in connection with the advent of a new decade, century or millennium?

In the early heyday of the concept of a new economy it was certainly costly to be stigmatized by financial markets as belonging to the old economy with its old industries and old, large companies run by old-fashioned managers. When the stock market bubble burst, it became even more costly to be associated with the new economy. Opinions were then frequently voiced that there was no such thing as a new economy and, if anything, there was "just" a gradual change in connection with emergent new technologies, infocom technologies (ICTs) in particular.

So, what is the factual basis underlying any sharp distinction between an old and a new economy and any claimed transition? No doubt, there has been a significant strengthening of capitalist economic systems around the globe after the downfall of the Soviet empire, and the spectacular and sustained resurgence of the US economy in the 1990s. Equally clearly, there has been a significant emergence of new ICTs (Internet, PCs, mobile communications, digital media etc.) in recent decades. On a grander time scale, economic growth and transformation have been spurred for centuries by cumulative new technologies, with technology-based firms increasingly becoming their main repositories, at the same time as markets have expanded through international economic integration.

In one sense it is now justified to speak of the emergence of a new economy, namely that intellectual capital rather than physical capital has come into dominance. This is a standard argument by proponents of the new economy. However, definitions and classifications of intellectual capital differ, and methods of measuring and accounting for intellectual capital are undeveloped to say the least.[2] Nevertheless, a number of indicators suggest that

[2] Here intellectual capital is taken to comprise all non-material (or intangible) resources that could be considered as capitalizable assets of an economic agent. Intellectual capital can be decomposed from the point of view of a firm into intellectual property (in patents, databases, trade secrets, trademarks etc.), relational capital (related to qualities such as trust in internal relations, constituting organizational capital, and qualities such as goodwill and reputation in external relations, constituting network capital) and human capital (related to competencies of various kinds).

there is a growing importance of different types of intellectual capital, and that there has been a cross-over to its dominance by some measures in the last few decades (see e.g. Granstrand 2000).

In addition, one can observe not only a growing share of intellectual capital in traditional firms, markets, products and professions, but also a growth of "pure" intellectual-capital firms, markets, products and professions.

Thus, there is in this sense a shift to a new economy (which consequently could be dubbed "intellectual capitalism"), in a mostly gradual and evolutionary manner. One main driving force behind this transformation or shift is technological change with the accumulation of new technologies in general, and in particular infocom technologies or ICTs. Through a number of key functionalities (enhancing e.g. codifiability of information and connectivity and excludability among agents), ICTs enable faster, more inexpensive and differentiated production and distribution of various old and new types of information that are of value to larger sets of users. ICTs also enable information to be commercially transactable at lower average transaction costs. Most importantly, ICTs enable more privatization of gains from production and distribution of information on a commercial basis, by embedding communication in "the world's largest machine" with possibilities to build electronic fences and locks, possibly reinforced by legislation (sometimes overly so). The traditionally recognized failure of information markets specifically due to appropriability problems thereby becomes mitigated.[3] Human communication then becomes linked through as well as to various machines, and thereby potentially becomes more easily commodified and commercialized through these various ICT functionalities. However, it would obviously be overly techno-centric to argue that ICTs in themselves are the only driving force behind the emergence of a new type of economy.

10.3 What Role has the Intellectual Property System Played?

If we are on our way in a new type of economy, what role has the IP system played and what is its role ahead? This must be left largely as an open question here, apart from a few observations. First, the IP system has historically been neither necessary nor sufficient for either technological or economic progress, e.g. in the first industrial revolutions (see Granstrand 2000). This is hardly a surprising statement but is nevertheless important to keep in

[3] For a classic work on market failures in information markets, see Arrow (1962), giving three general reasons for failure – uncertainty, indivisibility and inappropriability. For further interpretations, see e.g. Cheung (1986), Winter (1989), and David (1993).

mind, especially since technological progress is increasingly seen as necessary for economic progress.

Second, the IP system does not appear to have significantly fostered the specific technological progress in form of ICTs, at least not in their early stages up till the pro-patent era emerging in the 1980s. A few examples will illustrate. The transistor was patented at Bell Labs but licensed out generously, and the subsequent emergence of the semiconductor industry was significantly spurred by public procurement and a lax IP regime (Mowery 1996). The same could be said about the emergence of Internet under DARPA. The software industry also emerged under a lax IP regime (Samuelson 1993). The telecom industry was largely nationally monopolized until the 1980s and 1990s, with little interest in IPR. Mobile telephony also emerged until the late 1980s under a lax IP regime (Granstrand 2000). Thus, although ICTs have contributed significantly to the emergence of a new type of economy, IPRs do not appear to have contributed significantly to the emergence of ICTs, at least not until the 1980s. In fact it may even be argued that lax IP regimes were instrumental for the emergence of several ICT industries.[4]

Third, the comparatively rapid emergence of a much stronger IP regime since the 1980s (the pro-patent era) has by and large been concomitant with the much grander emergence of the new type of economy (intellectual capitalism), the roots and trends of which stretch much further back in history. The strengthening of the IP regime may very well have strengthened some features of intellectual capitalism and speeded up the development of some of its components in the last decade or so, but with our limited knowledge at present about the feedback structure involved it is only safe to say that the pro-patent era was as much, if not more, a consequence of intellectual capitalism as a cause of it. In any case a strong IP regime is a feature of the new economy with a concomitant expansion of IPRs by volume, type, and value as well as strategic attention paid to them.

[4] History in general has plenty of examples of how pockets of open S&T have been instrumental for progress, at least temporarily. These pockets or pools of open S&T may be open also to the general public by design (as with open standards or the open source movement) or by default. Commonly, they are closed or semi-closed with some kind of entry condition (e.g. granting back of improvements or agreeing not to take certain actions).

10.4 What are the Challenges to the IP System?

10.4.1 General fitness problems

The IPR system has certainly not lacked skceptics, critics and reformers during its many years of existence. Consequently there are a number of old challenges like international harmonization, rationalization of IPR organizations, high litigation costs, alternative dispute resolution schemes, monopolistic abuses, protectionist abuses, as well as effectiveness relative to alternative incentive schemes and means of appropriation. These challenges are still valid, and mostly increasingly so, as the IPR institution has become more important and expansive. Some of the challenges that seem to have emerged as particularly relevant or somewhat newly recognized will be focused upon here. No attempt is made to be exhaustive. Most likely there are further challenges ahead, still to be recognized, in the ongoing fitting of the IP system to the evolving new type of economy.

The rapidly changing character of technology easily creates misfits between technology and the more inert legal framework designed to foster it. This could be called the *IP-technology fitness problem*. It is an inherent tendency that changes in legislation lag behind changes in technology, legislation being by tradition and design more reactive and reluctant to *sui generis* approaches than pro-active and anticipatory when facing uncertainty. However, lags between new technologies and new legislation created to adequately deal with these technologies tend to become more common and conceivably more costly to society. Contemporary technology is advancing fast, perhaps faster than ever in absolute terms, with an entire array of new technologies emerging and combining into even newer ones. Some of these new technologies challenge fundamental concepts in the intellectual property system, and it is not clear at the outset whether and how they could be given intellectual property protection under the prevailing IP regime. Well-known examples are software and biotechnology. Less well known examples involve new surgical methods, new teaching methods, new business methods or even new athletic techniques.[5] As the cost and/or prospective value of new

[5] New business methods have become patentable in the US after a court decision in the so-called State Street Bank case in 1998. By now, prominent examples are Amazon's patent ("Internet-based Customer Referral System", issued as US Patent on Feb. 22, 2000) and Priceline's generic auction patent ("Method and Apparatus for a Cryptographically Assisted Commercial Network System Designed to Facilitate Buyer-Driven Conditional Purchase Offers", issued as US Patent on Aug. 11, 1998). This type of patents has stirred up a debate about what should be patentable subject matter, standards of patentability, proper scope of

technologies and inventions increase, the push for and disputes over IP protection will increase.

A related challenge is how to make the IP system better fit different industries or, more generally, how to make the IP system better fit the need to stimulate production and distribution of information and knowledge in different industries in light of available incentive systems and appropriability conditions. This could be called the *IP-industry fitness problem*. Tailoring of the IP system to fit the situation in different industries has been suggested from time to time (see Thurow 1997 for a recent example). A certain amount of tailoring does occur (e.g. through confinement of patentable subject matter, prolonged patent protection term in pharmaceuticals or industry-specific adjustment of patent scope plus through other IPRs) but has traditionally been done only to a limited extent. However, such tailoring is costly and, despite the fact that misfits are costly as well, significant industry tailoring is likely to be difficult and exceedingly costly. This is so primarily because of the complex and dynamic nature of the increasingly many-to-many correspondence between products and technologies, that is, the emergence of multi-product (generic, general purpose) technologies and multi-technology products (see below).

Another separate but related fitness problem is the *IP-nation fitness problem*. Needless to say, nations have widely different mixes of technologies and industries, and widely different needs for new technologies and industrial developments. As IP systems are tailored to stimulate innovations and their diffusion for the world as a whole, based on criteria not especially catering to nation-specifics (such as the novelty criteria), fitness problems are to be expected. International harmonization (for which there is also a need) aggravates this fitness problem when taken in a narrow sense of equalization, especially when primarily tailored ready-made to fit nations that are advanced or in other ways unrepresentative of the world spectrum. There is, of course, much to be said about this fitness problem, a problem that has been well recognized and debated over the centuries, but recently magnified and heated up with the advent of the pro-IP era, especially in connection with TRIPs and WTO. (See e.g. Mansfield and Mansfield 2000, and Maskus 2000.)

patent protection and proper cost/benefits of patent office operations (see e.g. Merges 1999). Some critics claim that such obvious but generic inventions, requiring almost no R&D, could jeopardize the development of e-commerce.

A fourth challenge is how to interface or harmonize different IP regimes. Different co-existing IP regimes have evolved over time, linked to different sectors of society and their institutions, organizations, norms, etc. Science and universities constitute one, technology and industry another, military and government another, culture and artists a fourth, and religion and churches perhaps also a fifth. These IP regimes are partly overlapping and interdependent, of course, but increasingly so as technology and economic concerns continue to penetrate modern societies. Thus, clashes between IP regimes occur more frequently, e.g. in industry/university collaborations. Pressures thereby arise to align different IP regimes with each other, and to find regime designs that are in some sense preferable on the whole. In this context one may note that universities, US ones in particular, have become economic institutions (Rosenberg and Nelson 1994, Nelson 1996 and Rosenberg 2000). Many public knowledge institutions in general have also become more business and proprietary knowledge-oriented, offering yet another indication of intellectual capitalism.

In this ambience the question must be asked whether there exists a single superior IP regime or a superior mix of complementary IP regimes for different economic and quasi-economic sectors of society. These sectors tend to gather creative and knowledge-producing people with different motivation structures and different propensities to be incentivized by standard utilitarian-based IPRs. This could be called the *IP-regime fitness problem*.

Finally, an *IP-criteria fitness problem* could be added as a separate category, even if it is closely related to the other categories. Tailoring of IPRs to fit their purposes could be done with regard to the whole set of IPRs as well as with regard to each IPR type, as to its scope of protection and criteria for being granted. While various IPR types are tailored very differently, each IPR type is fairly ready-made in terms of scope and granting criteria with fairly few variants. Rights-granting criteria easily create fitness problems as they are applied in a binary way, with a causal link to rights purposes easily becoming inadequate as the criteria set is stretched, mended, amended and adjusted to changing circumstances as time goes by.[6] This will be illustrated in the next section.

[6] Attempts to increase fitness in some area may also give rise to fitness problems in other areas, as unintended consequences might accrue. See Kash and Kingston (2001) for an illustrative view of this.

10.4.2 The IP-criteria fitness problem

As registered IPRs are granted based on a set of criteria, specific for each type of IPR, this set could become more or less fitting to the purpose of the IPR type in question as the criteria evolve and are applied over time. Patent rights provide a good example.

Patent rights are commonly granted for inventions on the basis of criteria regarding disclosure, unity of invention, usefulness or industrial applicability, novelties and inventive step or non-obviousness of inventions made in areas of patentable subject matter.[7]

A first question to ask is how well this set of criteria serves the purpose of the patent system, i.e. to stimulate innovation and diffusion when underinvestment could be expected. The criteria have been adopted fairly much the same across countries, but legal codes and administrative practices have evolved over long periods of time, while the nature of investment in innovation and diffusion has changed considerably, so some misfits could be expected. For example, in Europe patentable subject matter should have a "technical character". This limitation has a questionable basis in economic terms, since underinvestment in innovation and diffusion is not always present in or limited to technical fields. Besides, the boundary between what is technical and non-technical is fuzzy and changing. With this limitation the debate about patentability of e.g. software, allegedly lacking technical character, becomes too limited and distanced from economic considerations.

A second question is how to operationalize the different criteria and how to determine what level should be used for assessing whether they are fulfilled or not. For example, it is likely that, at some relaxed levels, a large volume of enforceable patents, dispersed among competing agents in an area, will become counterproductive due to potential or factual transaction costs and hold-up problems.[8] The area will simply be congested with rights, a danger that has been recognized since long ago. On the other hand, too hard-to-get patents in an area will reduce incentives to innovate to suboptimal levels, given that there is an underinvestment in the area in the absence of patent rights.

The *volume of patent applications* increases at the applicant end with (a) the availability of technological opportunities perceived as profitable by

---

[7] These are the common major criteria. There are a few additional ones, usually of minor importance or subsumable under the major criteria and therefore left aside here (e.g. regarding reproducibility and non-obscenity).

[8] At one extreme, patents could just be granted without examination, as has been the case in some countries. However, weak enforcement of such rights mitigates transaction costs.

(mostly overoptimistic) agents; (b) R&D investments and their yield or productivity; and (c) the patenting propensity by inventors and companies (influenced in turn by patenting costs and benefits, alternatives to patents, and habits). The *volume of patents granted* for a given quantity and quality of patent applications increases with (a) expansion of patentable subject matter; (b) lowering of patentability requirements such as inventive steps; and (c) shortening of processing time (temporarily). The *number of patent-blocking encounters* for a given volume of (maintained) patents then increases with (a) the dispersion of patents in a business area among a number of agents; (b) the number of patents needed to market a new product or process; (c) the number of new product and processes marketed; and (d) the scope of the patents. The *transaction cost* for buying a given blocking patent license then increases with (a) uncertainty about the scope of validity of the patent; (b) unavailability of substitute suppliers; and (c) information asymmetries and opportunistic behavior. In addition, the price of the license increases when the seller is a direct competitor and when the seller has a hold-up position.

Suppose we would like to redesign the criteria for patentability in order to reduce the sheer number of granted patents, so as to reduce transaction costs without reducing R&D incentives and payoffs too much. Then one can first observe that the set of criteria for patentability has a missing link to the purpose of the patent system. There is no explicit requirement to show that R&D efforts have been spent or are needed.[9] There is some, but not sufficient, correlation between the need for R&D investments and what are traditionally defined as areas of patentable subject matter, i.e. typical areas in engineering departments in industry and in technical universities. Thus, requiring in some sense substantial R&D efforts for patentability is one possibility.

Turning from missing to existing criteria, the *disclosure criterion* is probably a fairly functional criterion, although it has to be qualified as to how much should be disclosed ('enabling disclosure') and when.[10] The *unity of invention criterion* is somewhat fuzzy but probably not giving rise to severe dysfunctional effects.[11]

---

[9] This is in some contrast with the criteria for database protection in Europe.

[10] The timing of disclosure is highly important for its functionality. This brings in issues regarding consequences of late disclosure, e.g. in the form of so called submarine patents, generally seen as dysfunctional (see Granstrand 2000), and consequences of early disclosure, e.g. in a grace period. (For a good recent review of the grace period issue, see Straus 2001.)

[11] Apart from being a natural requirement regarding coherence in some sense, the criterion has been motivated by the need to prevent applicants from lumping applications together in order to lower application fees.

The *usefulness criterion* is the only criterion with some explicit economic bearing, but this criterion has been applied in a very weak sense if at all, though perhaps strengthened recently (Domeij 2000).

The *novelty requirement* induces a race and as such may have mixed effects. It is not clear that the best trade-off is when the winner takes all as in the patent system. Besides, novelty as well as usefulness is a matter of degrees and, without a qualification, congestion of rights easily obtains.

In fact the *inventive step requirement* was formed in Europe as a qualification or derivative of the novelty requirement, since novelty alone was perceived to allow uneconomically many patents to be granted. A criterion of this sort seems necessary, since taking it away would require some instrument for filtering out inventions with little usefulness or novelty that would anyway likely be generated without patent rights and would generate excessive transaction costs with patent rights.[12]

10.4.3 The IP assembly problem

At the level of innovations rather than at aggregate technology and industry level, a most challenging problem for the intellectual property system is what can be called the *IP assembly problem*. This problem constitutes an important source of transaction costs. Different technologies, old and new ones, have always been interacting with each other in complex and interdependent ways. However, this is increasingly so in the new economy with its increasing dominance of information (being highly recombinant and cumulative) and innovations (being new combinations of things in the economy). Opportunities to combine various technological advances can simply grow very fast. As a result, new products and services become not only increasingly based on new technologies, but increasingly based on many different technologies. That is, products and services become more multi-technological, or "mul-tech" for short, which is different from becoming "hi-tech" in the sense of using some specific advanced, new technology. At the same time more generic technologies appear, so in this sense technologies become also more multi-product or multi-process. All in all, the cross-links between new products and technologies proliferate. This means that patents and businesses become more cross-linked and interdependent, with each new

[12]The inventive step or non-obviousness requirement is in fact a fairly new one in the US. For its assessment in the US, see e.g. Adelman *et al*. (1998); in Europe, Cornish (1981), Domeij (2000) and Knesch (1994); and in Japan, JPO (1995) and JPO (2000). For the history of the inventive step requirement in various countries in North America and Europe, see Barton (2001), Beier (1985), Bochnovic (1982) and Ullrich (1977).

business becoming reliant on an increasing number of inventions and patents, and more new inventions and patents having an impact on a wide range of businesses.

Moreover, the sources of new technologies proliferate as more firms and nations invest in R&D, and firms also increasingly internationalize their sourcing and exploitation of new technologies. Thereby the number of firm-to-firm encounters increases as both factor input and product output markets become more global. The IP assembly problem is further aggravated by the recent trend in some fields, notably biotechnology, to grant patent and other IPRs (e.g. database protection) to research tools, i.e. inputs to the R&D process itself rather than to the downstream production process.

Thus, in a new technology as well as in a product area, there will be not only more agents on average, but increasingly interdependent agents in a mixture of cooperation and competition ("coopetition" or "competeration"). Technology trade, e.g. through licensing and cross-licensing, then becomes increasingly necessary. This is because IPRs to sustain a business become increasingly fragmented among players who are ready to enforce or otherwise exploit their rights, thereby creating transaction costs and possibly dynamic efficiency losses through a web of hold-up problems.[13]

The factors mentioned so far put increasing demands on the well-functioning of technology markets, which at the same time is actually facilitated by the IP system by design. What further complicates the functioning of technology markets is strategic firm behavior in using and abusing the intellectual property system. Some inventors and small firms without manufacturing act as "patent extortionists". Large corporations aggressively build up patent portfolios and employ various patent strategies. Moreover, they combine them with various other IPRs into a kind of multi-protection strategy, and thereby build up bargaining and retaliation power. IP management skills also develop in general. As IP-based bargaining powers are accumulated in industry, asymmetries in bargaining powers become more likely to appear – between new and old firms, between small and large firms, as well as between companies adapted to strong and weak intellectual property regimes in different sectors and countries. These asymmetries in bargaining powers create increasing risks in innovation and entrepreneurship, risks that become more and more difficult to absorb, especially for small manufacturing firms. (New kinds of insurances are being tried but with expensive premiums.)

[13] This would then be analogous to the anti-commons problem described in Heller (1998) with department stores in Moscow being unable to assemble all the necessary rights for their operations.

Thus, the intellectual property system may slow down, misdirect or hold up innovation and diffusion, although not necessarily discouraging all R&D investments. Consequently, there is, as there always has been, a mixed verdict over whether the intellectual property system promotes technological innovation and diffusion, but perhaps the doubts in the mix are increasing, not least due to increasing transaction costs. The pendulum continues to swing between trust in and suspicion of the intellectual property system.

10.5 What Remedies are there to the IP Assembly Problem?

There are a number of possible remedies to the IP assembly problem that can be considered, such as (a) more effective filtering out of insignificant patents through raising the standards of non-obviousness and usefulness, reducing the patent scope, steepening the renewal fee schedule, and implementing more efficient patent examination procedures; (b) using a two-tier structure with patents and utility models; (c) improving mechanisms for technology markets and transfer, including more liberal attitudes towards patent pooling and technology sharing, schemes for collecting rights and clearing-house procedures, cross-licensing and block-licensing incentives; (d) control of monopolistic abuse through reducing scope and length of protection and/or more consistent use of compulsory licensing; (e) reduction of legal uncertainty through faster and cheaper validation of rights and dispute resolution; and (f) facilitating the internalization of the necessary R&D through organizational integration, e.g. through consortia, joint ventures, mergers and acquisitions. Finally, IPRs could be abandoned altogether and replaced by alternative institutional arrangements (with their specific problems). For a summary of possible remedies, see Table 10-1, which also shows the available options or response strategies for a company encountering blocking patents.

Thus one possibility to mitigate the IP assembly problem is to raise the required level of invention (or level of non-obviousness) for patentability and to narrow the effective scope of a patent, especially in emerging technologies. This is done in order to decrease the hold-up problems and blocking power of many minor patents, as well as the blocking power of major patents in wide applications remote from their original ones.

Table 10-1: Remedies for the IP assembly problem (in the case of patents)

Society level	**Company level**
1. Filter out patents (reduce volumes) by raising: • Fees (base + renewal fees) • Examination quality • Standards of non-obviousness and usefulness	1. Invalidate blocking patents
2. Filter out patents by reducing: • Patent scope • Patentable subject matter • Length of patent protection	2. Invent around blocking patents
3. Improve technology markets by enabling: • Patent pooling and technology sharing • Block licensing and cross-licensing • Open source licensing • Rights collection and clearance	3. Obtain blocked technologies through acquisition, joint ventures or licensing
4. Reduce strategic behavior and abuse by: • Compulsory licensing • Injunctive reliefs	4. Cross-license or pool patents (e.g. through fencing-in or counterblocking in some other area)
5. Reduce legal uncertainty regarding: • Validation • Dispute resolution	5. Build up bargaining position, e.g. through partnering, purchasing power, patent power, and create credible threats
6. Enable integration and internalizing of interdependent R&D	6. Ignore or infringe blocking patents
7. Employ alternatives to private rights approaches, e.g. procurement contracts or public finance	7. Wait until blocking patents expire
	8. Stop relevant R&D and related commercial operations

The costs of patent processing and patent litigation, too, could be lowered in this way. Future patent flooding may also increase considerably when (rather than if) countries like China and India start patenting on a broad scale. Moreover, it is probable that something such as "computer-aided patenting" (or more broadly computer-aided invention) will come into play. That is, computers and expert systems can generate certain types of inventions (e.g. based on combinatorial alterations of inventive elements) or inventive inputs that form the basis for patent applications.[14] Such computer-

[14] This may appear a preposterous scenario to some, but is in fact quite feasible. Computer-generated musical compositions, art pieces, poems, visual patterns, logos, trade marks etc. are improving, as is the understanding of creative processes. So-called genetic algorithms also develop, with increasing capabilities to find optimal technical design configurations.

generated inventions are likely to be minor, at least in the beginning, and non-obviousness requirements must reasonably exclude such inventions. The impact on R&D and innovation efforts, as well as on diffusion of technical information, would probably not be significantly weakened in total by raising required levels of invention and limiting patent scope, but this issue requires further analysis.

An additional rationale for studying whether the inventive step requirement could and should be raised is that interviews with patent examiners and patent managers in leading countries (US, Japan and Europe) indicate that the inventive step requirement has become less stringently applied in recent years in some areas. Everything else equal, this would contribute to the proliferation and fragmentation of rights and their associated transaction costs, thereby magnifying the problems of fitting the IP system to the new type of economy.

10.6 Could the Inventive Step Requirement be Raised?

To assess whether the inventive step requirement is fulfilled, a suitably operationalized scale of non-obviousness and measurement procedures are needed.[15] Currently the scale in use is a binary one, with measurement commonly based on comparing a best reference document (BRD) (resulting from prior art search) and the claimed invention (CI), the comparison being done by a simulated synthetic person skilled in the art (PSA), who is specified in a certain way, and has to be distinguished from the inventor I. The mental step (effort) required for the PSA to move mentally from BRD to CI is then judged by the patent examiner (assisted by guidelines, experience and sometimes colleagues) as obvious or non-obvious or having a sufficient level of invention or inventive step, as the statutory language goes in some countries.[16] The final judgement is typically done after an exchange of requests,

[15] This assessment is related to assessing the size or radicalness of an invention, something that has been attempted by several authors over the years in different contexts (see e.g. Green *et al*. 1995). The presentation in this section is very stylized in an attempt to provide a formal representation as a first approximation, an attempt which by far does not capture all complexities involved in actual assessments of inventive step. The primary purpose here is to indicate how the requirement can be raised and lowered in principle.

[16] Note that the term 'non-obvious' refers explicitly to a mental process of the PSA, while 'inventive step' and some of its related present and earlier concepts in other languages (e.g. 'Fortschritt' and 'Erfindungshöhe' in German) refer more explicitly to the relation between the invention and prior art.

evidence, arguments and modifications between the examiner and the inventor, in which the burden of proof is shifted around, very briefly expressed.

There is now a technological distance TD between the BRD and the CI that has to be "travelled" by the PSA, as illustrated in Figure 10-1.[17] It seems impractical to adjust the amount of protection continually to the mental effort of the PSA (which however could be taken as an indicator of the amount of R&D and productivity needed in creative work). Retaining a binary outcome would still need a more structured scale if the inventive step or non-obviousness criterion is to be useful as a policy variable. At least a metric scale, if not normed, is then needed.[18] Lengths of inventive steps or amounts of non-obviousness then have to be defined in some way. One possibility is to have empirical reference points. Still, a theoretical foundation of measurement is needed. For this, a suitable operationalization of the concept of technological distance TD is needed for measuring the distance between the CI and the BRD, plus a mapping D of technological distances to mental efforts of a PSA, the latter suitably specified as a vector of attributes defining capabilities (knowledge, skills and equipment). An operationalization of the concept of mental efforts (or mental step) is then also needed.[19] The level of (mental) inventive step at time t (date of patent application) is then determined by $\widetilde{D} = E^{i}_{t+\Delta}\ \{D[TD(BRD_t(CI),CI),\ PSA_t]\}$, where $E^{i}_{t+\Delta}$ is the estimation of D made at time $t + \Delta$ by patent examiner i.[20] The amount of protection awarded could then be related to $\widetilde{D}$. Retaining the binary approach means specifying a level (vector) of possibly binary requirements (conditions) $\underline{D}$ and granting a patent to CI if, and only if, the corresponding $\widetilde{D} \geq \underline{D}$ (for some order relation). A change in the non-obviousness or inventive step re-

[17] For simplicity, the claim structure of the invention and the inherent nature of problems and solutions associated with the invention are not explicitly considered here. Also for simplicity reasons the CI, BRD, PSA and I are all represented just by points in a technology space, spanned by technology characteristics variables or invention attributes T_1 and T_2.

[18] A partially or even totally ordered scale of obviousness is not sufficient.

[19] Here one could conceive of a compound measure of technological and mental distances (or equivalently proximities) as well, but strictly speaking a measure of mental effort does not necessarily fulfill the triangle inequality as required for distance measures, basically saying that a detour can never be a short-cut. The latter is often the case in learning something new, however.

[20] It may appear an unnecessary exercise to use a symbolic expression like this. However, it captures in a compact way many of the elements involved (although without being complete here). Besides, any attempt to craft AI tools and decision-support systems for patent examination would involve symbolic representations. Note that features of the inventor are and should be irrelevant to the estimation.

quirement could then be achieved by changing $\underline{D}$ and/or re-specifying BRD, PSA, TD or D (e.g. by upgrading the PSA).

The question then is whether there is an optimal inventive step requirement $\underline{D}$ and how to characterize and find it.[21] The question is complicated by several facts. One is that inventive step is not assessed independently of other criteria. Thus considerations of novelty and usefulness and indirectly R&D efforts enter. This problem could be reduced to just considering $\underline{D}$ by assuming that there are monotonically increasing mappings from R&D efforts to technical performance T (i.e. the so-called effort curve in engineering) and from T to a utility function, plus a monotonically increasing mapping from T-differences to mental efforts. There is then a one-to-one relation between inventive step, R&D efforts and usefulness (utility). Equivalently we could then consider a minimal R&D effort $\underline{R}$ corresponding to $\underline{D}$.

Considerations of patent scope are more difficult to handle for two reasons. First and foremost, it is not clear how patent scope should be represented. Gilbert and Shapiro (1990) use the patentor's ability to raise price, Klemperer (1990) the impact on close product substitutes, and Lerner (1994) the number of side classifications of a patent. The probability of infringement or the invent-around costs could conceivably also be used. Second, there is an association of a certain patent scope with a higher inventive step (see Barton 1995), but its nature is not clear. However, with a given representation a reasonable specification could probably be made for joint optimization purposes (apart from the complexities involved).

Leaving the scope issue aside and optimizing with respect to $\underline{R}$ instead of $\underline{D}$, a formulation of the problem in the most simple case is given in the next section.

[21] There is a significant amount of literature on the optimal length and breadth (scope) of patent protection, as is well known, but to my knowledge almost nothing on the optimal inventive step (except for Hunt 1999). This is just a first attempt to deal with the question.

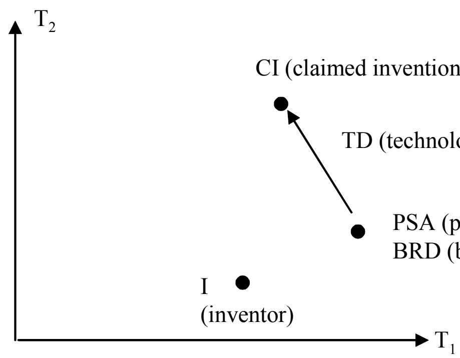


Figure 10-1: Positions in a technology-characteristics space for determining inventive step (non-obviousness) of an invention (represented by points for simplicity).

10.7 Optimal Inventive Step in the Simplest Case

Now let us turn to some formalized illustrations. The following variables and notation will be used below:

c	cost (variable cost c = c(q))
p	price
q	quantity sold at price p, i.e. the demand function is q(p) with an inverse demand function p(q)
r	discount rate
π	innovator's operating profit = $p \cdot q - c(q)$
R	total technology investment in in-house R&D and externally acquired technologies, with $\underline{R}$ = minimal R requirement; $\hat{R}$ = local, interior optimum; R_{opt} = glocal optimum; R_0 = value s.t. V = 0
B	transaction cost (incl. bargaining cost), with fixed and variable costs F and β
t	time (considered continuous)
V	total value discounted to time t = 0, i.e. present value; V^p producer surplus; V^c consumer surplus
W	Welfare = $V^p + V^c$

T	technical performance variable
L	length of period of patent protection
$\tilde{D}$	height of inventive step
$\underline{D}$	minimum inventive step as required for patentability
K, k, α	constants

All cash flow variables are functions of time corresponding to instantaneous cash-flow intensities, i.e. cash flows per unit of time.

The innovator's total discounted value V derived from a new product or process invention is then (with investments in production and marketing included in R):

$$V = \int_0^\infty \pi(t)e^{-rt}dt - \int_0^\infty R(t)e^{-rt}dt = V^p$$

To be able to aid R&D investment decisions in principle, the dependence of π upon R must be specified. This could be done by introducing a function T=T(R) (sometimes called the effort curve in engineering), a customer utility function U=U(T) and a T-specific cost function c=c(q,T). Considerations of T could also be bypassed by linking c directly to R through specifying c=c(q,R) instead.[22] One approach (Nordhaus 1969) is to introduce an "invention possibility function" for constant unit costs c (i.e. c(q) = c·q), being reduced from c_0 to c_1 by a cost-reducing patented invention derived from spending total R on cost-reducing R&D, according to:[23]

$$c_1 = c_0(1 - kR^\alpha) \;\; ; \; k > 0, \alpha \in (0,1), R \geq 0, c_0 > 0, c_1 > 0$$

Assume that the invention is sufficiently minor so that the pre-invention price $p = c_0$ and output q_0 remain the same until competition after patent expiration forces the price to fall to post-invention cost c_1 (assumed to occur immediately) with a corresponding expansion of output from $q(c_0) = q_0$ to $q(c_1) = q_1$. Then in the Nordhaus type of model, operating profit $\pi = 0$ when there is no patent protection and $\pi(t) = q_0(c_0\text{-}c_1)$ during patent protection.

---

22 Here, as is common, the symbols for variables and functions are used with differing connotations. Thus, the function c is not the same in the cases c(q), c(q,T), and c(q,R), but no misunderstandings are likely to derive from this convenient practice.

23 An alternative specification is $c_1 = c_o\,(1\text{-}e^{-\alpha R})$. Compare the specification in Wright (1983).

Inserting c_1 and assuming that q(p) is linear in p and that all investment expenditures occur at t=0 gives:

$V^p = q_0 c_0 k R^\alpha (1 - e^{-rL})/r - R = KR^\alpha - R$

The benefit to society of this invention then equals the producer surplus V^p, plus the additional consumer surplus V^c generated after patent expiration by the cost savings from the invention. Then (assuming the same discount rate r for consumers and producers):

$$V^c = \int_L^\infty q_0 (c_0 - c_1) e^{-rt} dt + \int_L^\infty \left(\int_{q_0}^{q_1} (p(q) - c_1) dq \right) e^{-rt} dt$$

The second term then corresponds to the dead-weight loss during patent protection being turned into consumer surplus after patent expiration when competition forces the price down to c_1. Now for q(p) linear in p:

$$V^c = (c_0 - c_1)(q_0 + q_1) e^{-rL} / 2r = (k_1 R^\alpha + k_2 R^{2\alpha}) e^{-rL}$$

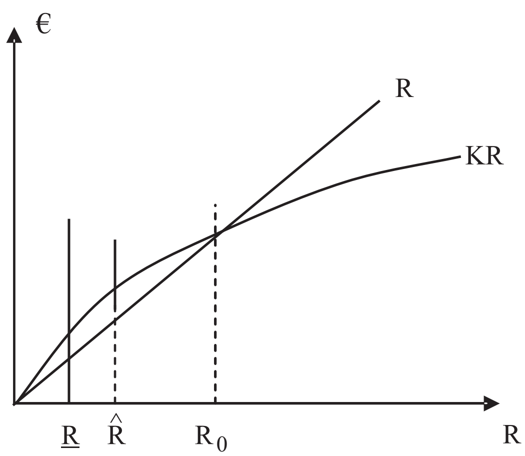


Figure 10-2: Optimal amount of R&D = $\hat{R}$ and minimal inventive step requirement, corresponding to $\underline{R}$, in the simple case of minor cost-reducing invention with linear demand.

Now, in order to simplify, include all productive investments in R, including costs for externally acquired technologies. Let transaction costs B that are borne by the company have two components, one fixed transaction cost F and one variable transaction cost ß(R), varying with R, i.e. B=F+β(R), as long as R > 0. Let B = 0 for R = 0. B includes patent fees and expected litigation costs as well as the company's bargaining costs for externally acquired technologies. If a certain share of new technologies needed by the company is externally acquired, and bargaining costs are proportional to the cost of these, we can let variable transaction costs be proportional to R as an approximation. That is, we can write β(R) = βR, where β is now a constant.

Thus, we have:

$$V^p = KR^\alpha - (1+\beta)R - F \qquad \text{for } R \geq \underline{R} \geq 0$$

$$V^c = (k_1 R^\alpha + k_2 R^{2\alpha})\, e^{-rL}$$

First, we study how β and F impact the optimal $R = \hat{R}$ and the maximal $V^p(\hat{R})$, and then how the inventive step requirement impacts β and F. Finally we look at impacts on V^c and total welfare $W = V^p + V^c$.

Differentiating V^p with respect to R gives an interior $\hat{R}$ as:

$$\hat{R} = (K\alpha/(1+\beta))^{\frac{1}{1-\alpha}} \qquad \text{and}$$

$$V^p(\hat{R}) = \hat{R}\,(1+\beta)(\alpha^{-1}-1) - F = (K\alpha)^{\frac{1}{1-\alpha}}\,(1+\beta)^{\frac{\alpha}{\alpha-1}}(\alpha^{-1}-1) - F$$

Since $d^2V^p/dR^2 < 0$, $\hat{R}$ is optimal if $\hat{R} > \underline{R}$ and $V^p(\hat{R}) > 0$.

If $V^p(\hat{R}) \leq 0$ then R = 0 is optimal.

For $\hat{R} < \underline{R}$, $\underline{R}$ is optimal as long as $V^p(\underline{R}) > 0$, that is, the inventive step requirement constraint is binding. If $V^p(\underline{R}) \leq 0$ when $\underline{R}$ is binding, then R = 0 is optimal.

Now we can see that in the case $\hat{R} > \underline{R}$ and $V^p(\hat{R}) > 0$, the following holds (see also Figure 10-2):

1) With no fixed but only variable transaction costs (i.e. F = 0, β>0), an increase in these through increasing marginal transaction cost β decreases $\hat{R}$ to $\underline{R}$ and V^p to zero.
2) With no variable but only fixed transaction costs (i.e. F > 0, β = 0) an increase in these through increasing F leaves $\hat{R}$ unchanged until V^p is decreased thereby to zero.

3) With both fixed and variable transaction costs increasing, V^p is decreasing for two reasons – partly since $\hat{R}$ is decreasing, and partly since V^p decreases for any fixed R.

Now what happens if we increase the minimum inventive step (patent height) requirement? The invention possibility function translates this requirement into a minimum technology investment requirement $\underline{R}$, so we can consider increases in $\underline{R}$ instead. From the above and Figure 10-2 we then see that if $\underline{R}$ increases from zero without increasing any transaction costs and $V^p(\hat{R}) > 0$, nothing happens until $\underline{R}$ increases beyond $\hat{R}$, in which case $\underline{R}$ becomes optimal and a loss in producer surplus is incurred since $V^p(\underline{R}) < V^p(\hat{R})$. Eventually $\underline{R}$ increases beyond R_0 in which case $V^p(\underline{R})$ becomes negative and R = 0 is optimal. Thus, the minimum inventive step requirement may become so high that it does not pay off on the whole to invest in new technologies, which require patent protection to pay off at all. Figure 10-3 depicts this case.

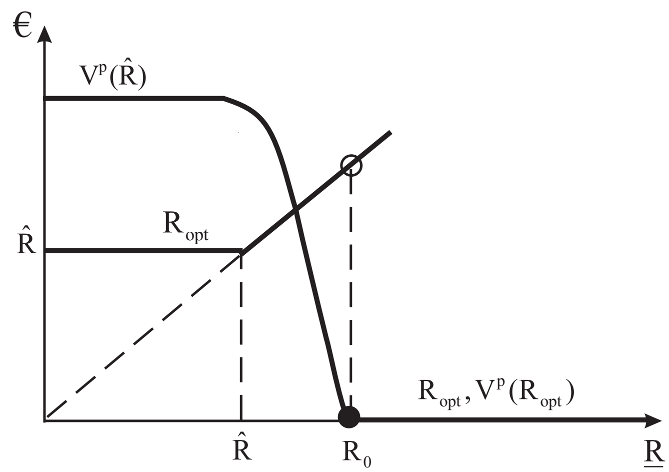


Figure 10-3: Impact of inventive step requirement $\underline{R}$ on technology investments R and producer surplus V^p.

However, changes in $\underline{R}$ will affect transaction costs. Transaction costs include patent fees and expected litigation costs.[24] With higher $\underline{R}$ there will be

[24] The classification of patent fees as transaction costs is not essential for the analysis here.

fewer possible patents per R&D dollar and fewer possible patent litigations, everything else equal, so these cost elements in B can be assumed to decrease in $\underline{R}$. Bargaining costs depend on the number of patent licenses acquired and this number goes down as $\underline{R}$ is increased, assuming that external technology acquisition cost is unchanged and that the transaction cost per license is constant. (The latter assumption could be replaced by the assumption that variable transaction costs depend only on the total acquisition cost (total in-licensing cost), not on the number of licenses, in which case $\underline{R}$ does not affect variable transaction costs, just fixed ones. In this case F decreases in $\underline{R}$, while β is unchanged, so $\hat{R}$ is unchanged within the range ($\underline{R}$, R_0) while V^p is increasing and so is R_0.)

If $\underline{R}$ is lowered, more patents can result per R&D dollar. Of course firms have the option to apply for patents with a higher than minimal inventive step requirement, or they can try to lump applications together. However, the unity of invention requirement has to be fulfilled. Besides, firms often use a patent strategy to apply whenever possible. (This strategy can be further analyzed in itself, of course.)

With more patents on average, the firms will have to pay more patent fees per R&D dollar and also higher bargaining costs for acquired technologies, being covered by more patents. Thus one can assume that variable transaction costs β(R) are proportional to $R/\underline{R}$ at least as an approximation, i.e. $\beta(R) = bR/\underline{R}$, where b is a positive constant.[25] Thus, in the case where both fixed and variable transaction costs decrease when $\underline{R}$ increases, V^p increases as long as it is positive and so does $\hat{R}$ and then optimal $R = \underline{R}$ when $\underline{R}$ is binding. Again for sufficiently large $\underline{R}$, technology investments do not pay off and optimal R is zero, together with V^p.

With this specification we have for the company's problem:

$$V^p(R, \underline{R}) = KR^{\alpha} - (1 + b/\underline{R})R - F, \text{ for } R > \underline{R}$$

$$\hat{R}(\underline{R}) = ((1 + b/\underline{R})/K\alpha)^{\frac{1}{\alpha-1}}; \qquad \lim_{\underline{R}\downarrow 0} \hat{R}(\underline{R}) = 0^+$$

$$V^p(\hat{R}(\underline{R}), \underline{R}) = \hat{R}(1 + b/\underline{R})(\alpha^{-1} - 1) - F = (K\alpha)^{\frac{1}{1-\alpha}}(1 + b/\underline{R})^{\frac{\alpha}{\alpha-1}}(\alpha^{-1} - 1) - F = \overline{V}^p(\underline{R})$$

[25] Here one may object that firms rationalize bargaining costs by cross-licensing, block-licensing and patent pooling, and also that the dispersion of relevant patents among firms is not proportional to the number of patents. Moreover, the propensity to acquire technologies

From this we see that $\hat{R}(\underline{R})$ and $\bar{V}^p(\underline{R})$ increase in $\underline{R}$ (with unrestricted R up to its optimal value without transaction costs), while the boundary $V^p(\underline{R},\underline{R})$ first increases, then decreases. In other words, increasing the minimal inventive step requirement will increase technology investments and producer surplus as long as the requirement is non-binding. If the require

$$V^p(\underline{R}, \underline{R}) = K\underline{R}^{\alpha} - \underline{R} - b - F$$

ment becomes binding, increasing it will increase technology investments, but producer surplus then increases only up to a point, beyond which it will decrease. If the requirement is further increased, producer surplus will eventually turn negative. Relaxing the requirement almost entirely, i.e. letting $\underline{R}$ approach zero, will push technology investments towards zero, since almost all resources then will be spent on patent fees, litigation and bargaining.[26]

Now for society's problem, we have for welfare W and $R \geq \underline{R}$:[27]

$$W(R, \underline{R}) = V^p + V^c = KR^{\alpha} - (1 + b/\underline{R})R - F + (k_1R^{\alpha} + k_2R^{2\alpha})\, e^{-rL}$$

V^c increases in R for all R (also for major process innovations in the Nordhaus framework) and company-optimal R increases with $\underline{R}$ (until it drops to zero). Thus, an optimal $\underline{R}$ has to be found among values where V^p decreases, which only happens when $\underline{R}$ is binding.[28] Increasing $\underline{R}$ then serves to increase consumer surplus at the expense of producer surplus to the point where the net effect turns negative.

At some point V^p is reduced to zero and so is company-optimal R; thereby the whole welfare effect is lost. Thus, depending upon parameter values, there is an optimal $\underline{R}$ distinct from zero and not too small and not too large. However, for certain parameter values, giving high patent fees and other transaction costs (through high b) but low R&D productivity (through

externally may change as more minor inventions are being patented. These objections, however valid, can be ignored at this level of a first approximation.

26 Litigation outcomes in terms of damages are assumed to leave V^p and V^c unaffected.

27 A more formal and detailed discussion, also regarding major ("drastic") process innovations in the Nordhaus framework, must be left aside here.

28 Note that for increasing patent length L, V^p increases since company-optimal R also increases, which in turn increases V^c but at the same time V^c is also decreased by a discount factor e^{-rL}. Increasing patent scope, simply in terms of amount of business applications, will increase company-optimal R and, through this plus a direct effect, increase V^p but decrease V^c through more dead-weight losses, although offset to some extent by increased R, so the impact on V^c is mixed.

low k and α close to 1), technology investments do not pay off at all to the firm. This holds when the patent lifetime L is held constant. (Note that several constants above depend on L.) However, L is also a policy variable and joint maximization of W with respect to both $\underline{R}$ and L involves a much more complicated analysis.

The analysis above is based on a stylized Nordhaus type of model, which leaves much room for discussion. Still, the analysis indicates that even in a simple but plausible case, an optimal inventive step requirement exists. However, how to find it is obviously non-obvious, requiring novel and inventive approaches for practical applications.

10.8 "Evergreening" by Incumbents

The analysis of a cost-reducing invention in a Nordhaus type of framework indicated the existence of a socially optimal level of inventive step requirement distinct from zero in certain parameter ranges. Raising the requirement from suboptimal levels would increase the available producer surplus and incentivize companies to invest more in R&D and new technologies, without having to spend most resources on patent fees and other transaction costs.

But if so, are there any reasons for companies to object to raising the inventive step requirement from already low levels without going too high? The analysis gives no such reason, as long as the requirement is non-binding; and any such requirement being binding on total R seems to be far from current reality.

Here, however, the limitations of the analysis must be taken into consideration. Neither R&D and technology acquisition, nor patenting, are decided and carried out as one-shot affairs in firms but take place intertwined in a process over time.[29] In such a process it may very well pay off to patent wherever and whenever possible, as witnessed in some industries and technologies with strong early mover advantages, e.g. in new, generic technologies with much uncertainty. A sequence of "small" patents granted at various steps in a process may then be more profitable (gives more producer surplus) to a firm than a single "large" patent granted at some specific step in the process, although difficult to assess *ex ante*. In fact, if an incumbent already has strong market, technology and patent positions in a certain product gen-

[29] The important issue of the role of patents in sequential invention has recently gained considerable attention; see Scotchmer (1991) for a pioneering article. See also Hunt (1999) for a study of the role of the non-obviousness requirement in a formal model of sequential patent races.

eration, continued follow-up patenting of even small product and process improvements and new applications serves to perpetuate or “evergreen” these positions. In other words, the effective patent protection is prolonged from a continually renewed patent portfolio. A new entrant trying to enter at some stage in this process will face unfavourable asymmetries regarding invent-around costs, transaction costs and bargaining power. This is not to say that it is always easy to evergreen through follow-up patenting, or that incumbents have always been good at it (which many of them have not), but lowering inventive step requirements opens up possibilities for once-innovative incumbents to maintain market power.

To clarify further, suppose there is an incumbent who has a major product patent A being complemented over time by a number of minor B-patents for product improvements, various add-on features, process steps and applications, as in Figure 10-4.

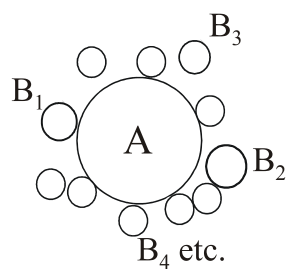


Figure 10-4: Major product patent A, being surrounded over time by follow-up minor B-patents B_1, B_2 etc.

Any new entrant wanting to invest in the A-technology when the A-patent expires also has to invest in or invent around B-patents. However, since product and process improvements derive much from learning by doing (learning by producing and using), the incumbent properly doing follow-up patenting is a likely source of many B-patents. The new entrant could of course invest more heavily than the incumbent in B-patents and succeed in doing so to the extent that the incumbent misses out in follow-up patenting.[30] However, the incumbent is well positioned on a set of learning curves and thereby likely to be relatively more efficient in this improvement type of

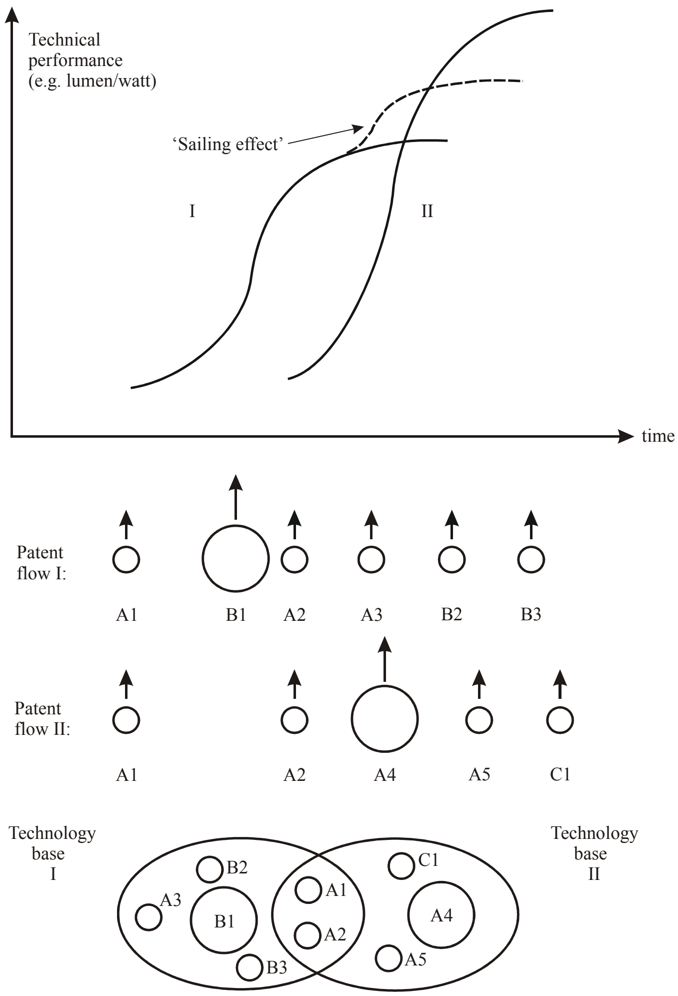
[30] A well-known alleged case of this is described in Spero (1990). DuPont’s patenting of nylon is a classic case of attempted but not entirely successful follow-up patenting (see Scherer 1980, p. 451).

R&D.[31] Thus, a low inventive step requirement allows the incumbent to sustain its market power through incremental innovations without having to recreate market power through more radical innovative R&D. The incumbent can then engage in patent protection of vertical as well as horizontal product differentiation. (The latter is also aided by large patent scope, as dealt with by Klemperer 1990.) Consequently the prospective new entrant has an incentive to focus on major innovations based on new technologies, and in particular with an initial business application not focused on by the incumbent. In fact, technology transitions and substitutions have historically typically provided new entry possibilities through major innovations, forming the basis for a new product generation. The issue for the incumbent is then how market, technology and patent positions in the old generation can be leveraged to the new generation. The subsequent issue here is how lowering the inventive step requirement is likely to influence incumbent leveraging and new entry possibilities. Figure 10-5 depicts the situation.

Figure 10-5 illustrates the patenting behavior of different competing companies A, B and C in two subsequent product generations, each with a technology base (portfolio of technologies), denoted I and II. These technologies may be partly competing (substituting), partly complementary. The shift from one product generation to the other typically involves a transition to new technologies, with some old technologies still being used and some rendered obsolete. The technical performance of the two product generations typically improves over time, as shown by the schematic S-curves in the figure.

The incumbent company B holds a major patent for technology base I and concentrates its R&D on further improvements in that area, while the incumbent company A builds patent positions in both technology bases. Company C is a new entrant in the area and focuses only on technology base II. The competing companies gain different patent shares in the different technology bases for the competing product generations, just as they may gain different market shares.

[31] Learning by using and learning by producing do not solely result in cost reductions through minor process innovations but also in improvements of product performance through minor product innovations (see in particular Sahal 1981 and Rosenberg 1982 for several illuminating empirical studies). The amount of these (mostly) minor innovations is correlated with cumulative production and stock of products in use (i.e. customer-installed base). Incumbents typically cumulate production experience (even in the case of outsourcing) as well as cumulating user experience through links to their customer base (see in particular von Hippel 1988).



Legend: A1, B2 etc = Company A's first patent in the area, Company B's second patent in the area etc.
I, II = Two technical performance curves, corresponding to technology base I and II, represented by two overlapping sets of technologies, being partly protected in technology space by two patent flows over time. The 'sailing effect' refers to improvements in old technical performance in response to threats from new technologies.
Circles denote scope of patents
Arrows denote patent granting dates

Figure 10-5. Patenting strategies in the case of competing or subsequent product generations.

Established companies with high market and patent shares for an established product generation are then well positioned to build up strong patent positions for a new product generation if this generation is technologically similar to the old one. Thus, incumbents are favoured by large overlaps (small technological distances) between the technology bases and patent portfolios for the old and new product generation, while new entrants are favoured by large degrees of substitution of old technologies for new ones. Both incumbents and new entrants are in addition favoured by technology overlaps offering possibilities to leverage their technologies in their other product areas, thereby reaping economies of scale as well as scope. This of course favors existing diversified companies rather than start-ups.

In addition, incumbent companies are better positioned to influence the rate and direction of technological changes. For example, an incumbent may introduce a "gap filler" which increases the technology overlap between the old and new generations, but at the possible expense of increased cannibalization. An incumbent may also influence standard-setting, as well as the adaptation to complementary products.

Now with the increasing multi-technology (multi-patent, multi-invention) nature of new products and processes and the emergence of generic technologies (multi-product technologies, general purpose technologies), the technology overlaps between different product generations as well as the technology overlaps between related but different product areas tend to grow over time. Increased possibilities to patent, enabled by lower inventive step requirements, then help companies in their technology leveraging. This circumstance helps incumbents to bridge over from an old to a new product generation. It also helps existing companies hitherto outside the product area to bridge over or diversify into the product area, using their technology base in other product areas as a technology platform or technology springboard. This is particularly so in situations of technological convergence or confluence (as in computers and communications), where the direction of technological changes may asymmetrically favor the diversifying new entrant.

The only category of firms that are not favored at all by this type of technology diversification at product level are young start-ups having to build up technology positions from scratch. But even if they are successful at that, they are most likely not able to become, let alone remain, technologically self-sufficient due to high R&D costs, boosted partly by invent-around costs from dense patent nets.[32] Thus, in their innovation process they will become

[32] The cost of inventing around a single patent (i.e. the cost of inventing a substitute technology for the patent) typically increases with its scope and inventive step but depends on many

dependent upon acquisitions of some technologies from incumbents and other existing companies and then subjected to transactional hazards and uncertainties. Their possibilities to make a significant entry depend on (a) the amount of technological substitutions imposing on incumbents costly and time-consuming scrapping of old technologies, and/or (b) technology management failures of incumbents (including omissions in patenting for evergreening) or (c) the possibilities to find alliances with firms trying to diversify or catch up. In the latter process, the start-up entrant could very well be acquired fully or find its role reduced to that of a technology sub-supplier, lacking complementary resources.

What has been described so far is, of course, not a direct and definite outcome of a low level of an inventive step requirement, but lowering this requirement is likely to contribute to such an outcome. Thus, an argument could be made that it is in the interest of incumbents to have a low inventive step requirement when sequential invention and patenting is considered.[33]

10.9 Alternatives to Patents

Finally, when addressing the optimality of the inventive step requirement for patent protection, one could address the optimality of patent protection for innovation and diffusion inducement as well. This is a large and important topic for which a new angle of analysis will be given here and then left to further research.

Competitive markets could be expected to fail to generate adequate private investments in R&D and new technologies due to problems of e.g. appropriating returns on costly new information (Arrow 1962). Information has the standard properties of public goods, with non-excludable, non-rival consumption, as well as low variable and high fixed costs in production, therefore justifying some form of public provision and finance.[34] David (1993) distinguishes three approaches to such public provision – property, procure-

---

more factors, e.g. its position in the technology landscape and the skills of the parties involved. The cost of inventing around several related patents could be expected to increase far more than proportionately to the number of patents to circumvent.

[33] It is not immediately clear that this is welfare-reducing. To the extent that cross-licensing among incumbents functions, the situation could be seen as a form of "oligopolistic prospecting" to which the prospect theory of Kitch (1977) could possibly be extended. However, the standard counterarguments to Kitch's theory (see e.g. Kaufer 1989) apply as well and with even greater strength. In addition, cannibalization and strategic gaming and defection among the incumbents weaken the coordination or governance of the prospecting.

[34] Information has specific characteristics in addition, such as being cumulative and interactive (in the words of David 1993), thus making information a specific type of public goods.

ment and patronage (e.g. public grants or prizes to research) in his terms.[35] All approaches are used in a mixed approach in reality, but it is not clear that an optimal solution would be a mix and it is far from clear how to find it.[36] However, information and knowledge are uncertain and highly heterogeneous and so are the conditions under which they are produced and diffused, justifying a combined variety of approaches to foster suitable conditions. The heterogeneity could be roughly characterized in terms of a few variables related to the public goods characteristics used to assess the various approaches. Thus, using the property approach means decentralizing decision-making about scarce resources to agents with unique access to localized information for proper decisions, and incentivizing them to exercise their capabilities by providing them with access to a share of the extra surpluses they then generate. The latter is done by allowing the property holder to charge prices higher than marginal cost in order to help cover fixed investment costs.

Such monopolistic pricing is a drawback of the property approach, as it incurs a certain loss of consumer surplus apart from a shift of some surplus from consumers to the producer. However, in order to assess the property approach, this drawback has to be compared with corresponding drawbacks of other approaches. If the right to exercise certain monopolistic pricing is seen as a decentralized right to tax consumers, it corresponds to the right to impose a targeted sales tax administered by private agents. The administrative cost (transaction cost) could then be fairly low in comparison with public forms of taxation, be they targeted (selective) or general.[37] Of course, taxes could be more than minimally distorting and over-taxation could occur, as it could with any form of tax. (Few people seem to disagree on this.) The real virtue as well as drawback of the property approach, I would argue, is its amenability to flexible decentralization as well as to over-decentralization in the sense that too many and costly agent interdependencies will arise, resulting in too high transaction costs, eventually high enough to outweigh incentive effects and other efficiency gains. In addition, recentralization is usually more difficult (costly) than decentralization.

---

[35] To these three approaches could be added a fourth – private prizes (e.g. the Nobel Prize).

[36] A pioneering article here is Wright (1983), showing when patents, prizes and contracts are optimal.

[37] Just to mention one comparable alternative, consider the popular use of R&D tax credits or tax deductions for stimulating innovation, based on the idea to subsidize R&D inputs through targeted cuts in general taxes. This tax arrangement has significant limitations and hardly qualifies as a minimally distorting tax arrangement (see Mansfield 1982). It could be modified of course, e.g. to cover commercial activities as well, not just R&D, but it will still be inherently limited (see Granstrand 1998).

This view of the property approach is actually an organization-theory view, in which the handing out of private property rights is comparable to handing out or decentralizing responsibilities and accountabilities in an organization, applying management by objectives, dealing with principal-agent and information asymmetry problems, intervening for conflict resolution and so on. There are many organizational principles, some of which are that decentralization should be aligned to the information structure, incentive structure and structure of interdependencies, and should not be carried so far that the management cost of coordination outweighs benefits, e.g. from entrepreneurial motivation and economies of scale from division of labor. The adoption of the so-called M-form of organization, that is, a form of organization of a firm being decentralized into product divisions, is a recognition of stronger interdependencies within product-related activities than e.g. within functionally related activities (i.e. activities within R&D, production and marketing functions). Handing out patent rights to product inventions is then comparable (but not equivalent) to adopting an M-form of decentralization. If, however, inter-product interdependencies become more costly to coordinate than intra-product ones, the M-form has been carried too far, with too many small interdependent divisions with internal transfer pricing problems (transaction costs), conflicts and costly higher management intervention (the organization's internal court system). Thus, using a property approach is largely a matter of how far decentralization should go along what organizational principles, in order not to let transaction costs outweigh innovative and efficiency gains by handing out too many small interdependent property rights.

Finally, a few remarks on the other approaches are in order. With very generic, uncertain and interactive information, especially interactive without being embodied, more efficient forms of financing and taxation than through the property approach are conceivable. This could for instance be the case for basic scientific research. With specific, more certain and embodied information, targeted public procurement financed by public taxes could be a proper approach. This could be and often is the case for large, infrastructural development projects. (See again Arrow 1962.) If fame and prestige, and expost financing, are relatively more important incentives for producing information, then prizes might be a proper approach. Behaviors like defensive publishing and open source licensing provide other cases for similar analysis. The point is that, in considering demand and supply conditions when judging the various approaches to provide for innovation and diffusion and how they can complement and substitute each other, one should also take into account: (a) the nature of information (knowledge), e.g. in terms of genericness, uncertainty and interactivity, and (b) the nature of conditions surrounding information production and diffusion, e.g. in terms of information asymmetries, incentives and financing needs, costs of taxation and other transactions, and licensing conditions.

10.10 Summary and Conclusions

A new type of economy has gradually emerged, much as a result of a long-running cumulation of once new and still valuable technologies, not least ICTs, which have dual roles as generating value and enabling the privatization of value. Intellectual capital – including human capital, relational capital and IPRs – has come to dominate in some sense over physical capital. A pro-patent or pro-IP era has rapidly emerged since the 1980s, embedded in the much slower emergence of a new type of economy, apparently more as a consequence than a cause to it. Reassessments of old problems associated with IPRs and identification of new or recently magnified ones are then called for. The chapter briefly describes the increasing problems of fitting the IP system to technology, industry and country specifics in the new economy, and then focuses specifically on the IP assembly problem, magnified by congestion of various IPRs. Thereby intellectual resources become fragmented ("the anti-commons problem"), thus magnifying transactional problems, calling for more attention paid to transaction cost analysis of IPRs in the new type of economy.

Transaction costs arise from many sources, e.g. emergence of generic (general purpose) technologies and multi-technology products and processes, combinatorial growth of new technologies, growth of IPRs by types, numbers and scope and strategic use of IPRs by firms. The chapter briefly describes a number of remedies to the IP assembly problem and the role of patentability criteria, especially the non-obviousness criterion or inventive step requirement. This requirement was once introduced as a qualification of the inventive novelty requirement in order to reduce the congestion of patents. Interviews with industry and patent examiners indicate that the requirement has become less stringently applied at places and/or that it could be too low for its purposes. The chapter makes a first attempt to analyze the optimality of the inventive step requirement, taking transaction costs into account. The analysis of a simple, single invention framework of the Nordhaus type shows the unique existence of an optimal minimal requirement level, below which too many resources are spent on bargaining and patent fees, and above which R&D investments become sub-optimal. A qualitative analysis of sequential invention in a dynamic framework supports the argument that a low inventive step requirement favors incumbents, who then can engage in “evergreening” more efficiently to sustain market power in connection with technological transitions. Such transitions otherwise offer a window of opportunity for new entrants, especially if they involve bridging large technological distances.

Further research is warranted on the macro- and micro-economic implications of this micro-legal issue, e.g. regarding the IP assembly problem and its remedies, transaction cost analysis of inventive step and related patentability criteria, and empirical research on patent examination procedures. At a higher level, further research is warranted on optimality conditions for various institutional arrangements for innovation, diffusion and S&T governance in general, patents included, taking total governance costs into account.

In summary, there are reasons to be concerned that transaction costs make patents sub-optimal in certain areas, e.g. in basic research and generic technologies, and that the inventive step requirement should be increased. This, however, also raises the questions whether patents, even with optimized patentability criteria, are optimal in comparison with other institutional arrangements, such as prizes or contracts.

10.11 Literature References

Adelman, M.J., Rader, R.R., Thomas, J.R. and Wegner, H.C. (1998). *Cases and Materials on Patent Law.* American Casebook series, West Group.

Arrow K. J. (1962). ‘Economic Welfare and the Allocation of Resources for Invention’. In *National Bureau of Economic Research*, 1962. *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton Univ. Press, pp.609–626.

Barton, J.H. (1995). ‘Patent scope in biotechnology’, *IIC*, *International Review of Industrial Property and Copyright Law*, 26 (5), pp. 605-618.

Barton, J.H. (2001). *Non-obviousness*, unpublished mimeo. Stanford University, Law School.

Beier, K-F. (1985). “Zur historischen Entwicklung des Erfordernisses der Erfindungshöhe”. *GRUR*, Heft 8, pp. 606-616.

Bochnovic, J. (1982). ‘The Inventive Step: its Evolution in Canada, the United Kingdom and the United States’, *Studies in Industrial Property and Copyright Law*, Vol. 5, Weinheim: Verlag Chemie.

Cheung, S.N.S. (1986). ‘Property Rights and Invention’. In Palmer (1986), pp. 5-18.

Cornish, W.R. (1981). *Intellectual Property: Patents, Copyrights, Trade Marks and Allied Rights.* London: Sweet & Maxwell.

David, P.A. (1993). ‘Intellectual Property Institutions and the Panda’s Thumb: Patents, Copyrights, and Trade Secrets in Economic Theory and History.’ In Wallerstein *et al*. (1993), pp. 19–61.

Domeij, B. (2000). *Pharmaceutical Patents in Europe.* The Hague: Kluwer Law International.

Gilbert, R. and C. Shapiro, (1990). 'Optimal patent length and breadth', *RAND Journal of Economics*, 21 (1), 106-112.

Granstrand, O. (1998). *R&D Tax Credits – Why they mostly do not work and how they perhaps could be designed to work.* CIM Working Paper No 1998:02, Chalmers University of Technology, Göteborg, Sweden.

Granstrand, O. (2000). *The Economics and Management of Intellectual Property*. Cheltenham, U.K: Edward Elgar Publishing Ltd.

Green, S.G., Gavin, M.B. and Aiman-Smith, L. (1995). 'Assessing a Multidimensional Measure of Radical Technological Innovation', *IEEE Transactions on Engineering Management,* 42 (3), pp. 203-214.

Heller, M.A. (1998). 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets', *Harvard Law Review*, 111 (621), pp. 622-686.

Hunt, R.M. (1999). *Nonobviousness and the incentive to innovate: An economic analysis of intellectual property reform.* Working Paper No 99-3, Research Dept., Federal Reserve Bank of Philadelphia. Philadelphia: Federal Reserve Bank.

JPO (1995). *Practices in Examination and Appeals under 1994-revised Patent Law in Japan*. AIPPI, Japan.

JPO (2000). *Examination Guidelines for Patent and Utility Model in Japan*. Guidelines, JPTO, Tokyo, Dec. 2000. (Downloadable also from JPTO homepage.)

Kash, D.E. and Kingston, W. (2001). 'Patents in a world of complex technologies', *Science and Public Policy*, 28 (1), February, pp. 11-22.

Kaufer, E. (1989). *The Economics of the Patent System.* New York, NY: Harwood Academic Publishers.

Kitch, E. (1977). 'The nature and function of the patent system', *Journal of Law and Economics*, 30 (1), pp. 265-290.

Klemperer, P. (1990). 'How broad should the scope of patent protection be?', *RAND Journal of Economics*, 21 (1), pp. 113-130.

Knesch, G. (1994) 'Assessing Inventive Step in Examination and Opposition Proceedings in the EPO', *EPI Information*, 3, pp. 95-99.

Lerner, J. (1994). 'The importance of patent scope: an empirical analysis', RAND *Journal of Economics,* 25 (2), pp. 319-333.

Mansfield, E. (1982). 'Tax Policy and Innovation', *Science*, 215 (12), pp. 1365-1371.

Mansfield, E.D. and Mansfield, E. (guest eds.) (2000). 'Intellectual Property Protection and Economic Development', *Int. Journal of Technology Management*, Special Issue, 19 (1/2).

Maskus, K.E. (2000). *Intellectual Property Rights in the Global Economy.* Institute for International Economics, Washington DC. August, 2000.

Merges, R.P. (1999). *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts*. Mimeo, UC Berkeley School of Law.

Mowery, D.C. (ed.). (1996). *The International Computer Software Industry. A comparative study of industry evolution and structure.* Oxford: Oxford University Press.

Nelson R.R., (1996). *The Sources of Economic Growth.* Cambridge, MA: Harvard Univ. Press.

Nordhaus, W.D. (1969). *Invention, Growth and Welfare.* Cambridge, MA: MIT Press.

Palmer, J. (1986) (ed.). 'The Economics of Patents and Copyrights', *Research in Law and Economics*, Vol. 8, Greenwich, CT: JAI Press Inc.

Rosenberg, N. (1982). *Inside the Black Box: Technology and Economics*. Cambridge, MA: Cambridge University Press.

Rosenberg N. and Nelson R.R. (1994). 'American universities and technical advance in industry', *Research Policy* 23, pp. 323–348.

Rosenberg, N. (2000). "American universities as endogenous institutions", published as Ch. 3, pp. 36-57, in Rosenberg, N. (2000). *Schumpeter and the Endogeneity of Technology*, Routledge, London.

Samuelson P. (1993). 'A Case Study on Computer Programs'. In Wallerstein *et al*. (1993), pp. 284-318.

Scherer, F.M. (1980). *Industrial Market Structure and Economic Performance.* Chicago, IL: Rand McNally.

Scotchmer, S. (1991). 'Standing on the shoulders of giants: cumulative research and the patent law', *Journal of Economic Perspectives,* 5 (1), pp. 29-41.

Sahal, D. (1981). *Patterns of Technological Innovation.* Reading, MA: Addison-Wesley Publ.

Spero, D.M. (1990). 'Patent Protection or Piracy – A CEO Views Japan', *Harvard Business Review*, 68 (5), pp. 58-67.

Straus, J. (2001). *Grace Period and the European and International Patent Law – Analysis of Key Legal and Socio-Economic Aspects.* IIC Studies in Industrial Property and Copyright Law, Vol. 20. München: Verlag C.H. Beck.

Thurow, L.C. (1997). 'Needed: a new system of intellectual property rights', *Harvard Business Review* Sept-Oct, pp. 95–103.

Ullrich, H. (1977). *Standards of Patentability for European Inventions – Should an Inventive Step Advance the Art?* IIC Studies in Industrial Property and Copyright Law, Vol. 1. München: Max-Planck-Institut.

Wallerstein, M.B., M.E. Mogee and R.A. Schoen (eds.) (1993). *Global Dimensions of Intellectual Property Rights in Science and Technology*. National Academy Press, Washington, D.C.

Weil, V. and J.W. Snapper (eds.) (1989). *Owning Scientific and Technical Information: Value and Ethical Issues*. New Brunswick and London: Rutgers University Press.

Winter, S.G. (1989). 'Patents in Complex Contexts: Incentives and Effectiveness'. In Weil and Snapper (1989), pp. 41-60.

von Hippel, E. (1988). *The Sources of Innovation.* New York, NY: Oxford University Press.

Wright, B.D. (1983). 'The economics of invention incentives: patents, prizes, and research contracts', *American Economic Review*, 73 (4), pp. 691-707.