P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-FM CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 17:23

Industrial Organization

Markets and Strategies

2nd edition

Paul Belleflamme
Université Catholique de Louvain

Martin Peitz
University of Mannheim

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107687899

First published 2010
Second edition 2015

Printed in 'country' by 'printer'

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN 978-1-107-06997-8 Hardback
ISBN 978-1-107-68789-9 Paperback

Additional resources for this publication at www.cambridge.org/belleflamme2

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

Part VII
R&D and intellectual property

Introduction to Part VII: R&D and intellectual property

Early March 2014, one could read in the press that Twitter had paid IBM $36 million for 900 patents in order to avoid a lawsuit (Wired.com, March 7), that a German company had won the EU's €2 million inducement prize for innovative vaccine technology (European Commission, March 10), that SanDisk had filed lawsuits against its competitor SK Hynix, alleging misappropriation of confidential trade secrets related to flash memory technology (*PCWorld*, March 13), that the US Supreme Court was considering making software ineligible for patent protection (*Forbes*, March 16), that Viacom and YouTube had settled a copyright violations battle out of court, seven years after their dispute began (*The New York Times*, March 18) and that Nissan and its French partner Renault were planning to integrate their research and development functions (*The Japan News*, March 19).

What do these events have in common, except that they occurred at about the same time? They all involve intangible assets (i.e., assets that do not have a physical or financial embodiment), which consist of human knowledge and ideas and to which a legal entitlement, called *intellectual property* (IP), is usually attached. Intangible assets of this kind become increasingly crucial in our economies. In some countries, the investment in intangible assets now matches or even exceeds investment in tangible assets (such as machinery, buildings and equipment). The causes have to be found in the growing importance of service industries, in globalization and in the fast development of information and communication technologies (ICTs).

A clear manifestation of this trend is that firms are increasingly seeking patents, attempting to extend their scope, granting more licences, litigating more and transforming their business models around intellectual property. At the same time, highly reliable open-source software is collectively produced by a decentralized crowd of developers who do not seek any immediate monetary compensation for their efforts; commercial software vendors are contributing to these projects and, even more surprisingly, they do sometimes initiate open-source projects by releasing part of their proprietary source code.

Similar trends are observed in the entertainment industries (music, cinema, videogames, etc.). For instance, on 18 March 2014, the International Federation of the Phonographic Industry (IFPI, which represents the recording industry worldwide) attacked Google for failing to address Internet piracy after revealing a $600m decline in worldwide music sales in 2013. In contrast, Creative Commons, a non-profit group aimed at carving out ways to share creative works, is expanding from the realm of copyright into patents and scientific publishing. Creative Commons has built a licensing system that allows content creators to decide which usage rights to their work to grant others.

These contrasting examples raise the following questions: *Why is intellectual property used and diffused in such contrasting ways? Who are the winners and who are the losers in these different situations? What is best for society as a whole?* These questions appear all the more important now that intellectual property stands at the heart of industrial and economic policy. For instance, the European Commission launched in 2010 the *Innovation Union* initiative, which, with grandiose words, aims to 'make Europe a world-class performer in science, revolutionise the way the public and private sectors work together, notably through Innovation Partnerships, and remove bottlenecks – create an internal market for skills, patents, venture capital, innovation procurement and standard setting to foster ideas being quickly implemented on the market'.[1] At national and regional levels, governments also take special measures to boost the economy through a substantial increase in the public funding of both fundamental and applied research, based on the observation that *innovation* is one of the main sources of economic growth.

The importance and prevalence of the previous issues call for careful analyses of the mechanisms and institutions governing the production, use and diffusion of intellectual property. In this respect, the economic approach appears fundamental as it focuses on markets, incentives and strategic interaction. In this part of the book, we aim to develop a rigorous economic analysis of a large set of issues surrounding intellectual property, R&D and innovation.

Part VII is concerned with activities generating information and knowledge. These activities suffer from a generic problem of appropriability, because they entail externalities, indivisibilities and uncertainty. As a result, investments in information and knowledge are quite different from other investments made by firms or individuals. In particular, when left alone, markets generally fail to provide the right incentives to make these investments. Institutions thus need to be put in place in order to improve the provision of information and knowledge. The most prominent of these institutions is the protection of intellectual property. In Chapter 19, we explain its economic rationale and describe two of its main forms: patents and copyright. We also examine a number of important issues related to the protection of IP. Which mix of length should patents and breadth have? What are the perverse effects of the patent system (duplication of efforts in patent races, 'blocking patents', patent filing for defensive reasons, etc.)? Is the patent system appropriate for cumulative innovations? How do licensing and patent pooling affect the optimal design of the IP system? What are the effects of digital technologies and the Internet on the protection of IP?

In Chapter 18, we examine the two-way relationship between market structure and the incentives for R&D, focusing first on the profit incentive to innovate and then considering the strategic incentives to innovate in situations where firms compete in R&D. We also compare the pros and cons of various forms of cooperative R&D relative to non-cooperative R&D for firms competing in a product market.

The two chapters differ in the line of analysis they pursue: Chapter 18 mostly addresses positive questions, while Chapter 19 focuses on normative questions and applications to specific industries. As a result, the two chapters also differ with respect to the balance between theory and cases: Chapter 18 is more 'theory-intensive', while Chapter 19 is more 'case-intensive'. In particular, Chapter 19 provides the reader with insights on R&D and intellectual property in specific industries, in particular pharmaceuticals, software and digital music.

What will you learn in this part of the book?

You will first enlarge the perspective of the previous parts by incorporating innovation into the picture. You will see how firms use innovation as a strategic weapon, either on their own or in cooperation with some of their competitors. You will also understand why firms' investments in R&D are often insufficient, but may sometimes be excessive from society's point of view. You will then learn how to organize your thoughts about a number of topical issues related to innovation policy. In particular, you will understand the role of the patent system, and you should be able to form your own opinion about the proposed reforms of this system.

[1] See European Commission (2013).

18 Innovation and R&D

In this chapter, our goal is to examine the interplay between market structure and innovation. This is clearly a two-way relationship: on the one hand, firms' incentives to invest in R&D depend on the structure of the product market they are acting in (i.e., on the number of rival firms and on the way they compete); on the other hand, firms are likely to use R&D to shape the structure of their market (e.g., by using R&D to increase their market share or to keep potential competition at bay). As the two effects are complex and intertwined, we simplify the analysis by assuming that firms can somehow appropriate the return from their R&D investments (we analyse how they actually manage to do so in the next chapter). We also break down the analysis into separate issues.

In Section 18.1, we assess how market structure affects the incentives for conducting R&D, which are measured by the profit increase that the innovator gains from the innovation. First, incentives to innovate are compared in the two market structures where strategic considerations are absent, namely monopoly and perfect competition. It is shown that the latter generates larger incentives to innovate than the former. Next, we extend the analysis to include strategic interaction by considering oligopolies. We reach an ambiguous result: a higher intensity of competition may increase or decrease the incentives to innovate depending on the initial starting point, the size of the innovation and the way competition is increased. Finally, we study the possibility for the innovator to obtain additional revenues by licensing its innovation to other firms, be it inside or outside its own industry.

In Section 18.2, we reverse our point of view by investigating how innovation may influence market structure. First, we reconsider a monopolist's incentives to innovate in situations where a competing firm threatens to enter the market. We show that the incumbent firm is often keener to invest in R&D than the entrant. Monopoly is thus likely to prevail over time, which indicates that innovation does indeed drive market structure. Second, we enrich the previous analysis by incorporating explicitly the time dimension. Indeed, a firm's main motivation when investing in R&D is often to be the first to come up with an innovation. In such a *race* to be the first innovator, firms have to decide about the timing of their R&D investments. In this dynamic framework, we reconsider the comparison between the R&D decisions of an incumbent and of an entrant. We also examine the possibility that such races might exacerbate incentives to invest in R&D in such a way that total investments exceed the socially desirable level.

In Section 18.3, we consider other strategic aspects of R&D investments that arise when firms anticipate the effects their R&D investments will have on the product market competition. Here, the questions of interest are the following. Do firms invest more or less when they recognize the strategic nature of their R&D decisions? Should firms be allowed to coordinate their decisions at the R&D stage? How do the answers to the previous two questions depend on the public-good nature of R&D?

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

Nature of technical progress

Before we enter into the detailed analyses, it is important to note that the three sections of this chapter differ not only in the issues they address but also in the assumptions they make about the nature of technical progress. As will be made explicit below, we assume in Section 18.1 that R&D investment determines (instantaneously and for sure) the size of the innovation and that only a single firm ends up using the innovation. In contrast, Section 18.2 assumes some form of 'tournament technical progress', where the timing of innovation is uncertain and depends on the R&D investments of all firms and where the size of the innovation is fixed. Finally, Section 18.3 resembles Section 18.1 in that the tournament aspect and uncertainty are left aside, but it differs in two major aspects: first, the size of the innovation depends on the intensity of the firm's R&D investment (and potentially on the other firms' investments as well) and second, firms have the simultaneous opportunity to achieve competing innovations.

This variety of settings may seem confusing at first glance. However, it is motivated by the fact that the innovation process itself differs widely across situations and industries. In this respect, it is useful to distinguish between two types of innovative environment.[2] On the one hand, when the creation of knowledge addresses a known need, it seems logical to assume that *ideas are common knowledge*. In other words, any good idea is likely to be had and implemented by someone else. In this environment, discovery is seen as the result of some exogenously given production function that transforms R&D inputs into an invention of a certain quality, or into a probability of finding a new product or process. This environment corresponds to what is assumed in Sections 18.2 and 18.3. On the other hand, there are also instances where the need had not been identified prior to someone thinking of the idea. It is then natural to assume that *ideas are scarce*, in the sense that there exists no substitute idea that would address the same economic need (or that each idea occurs to a single, random person). This environment fits more the framework assumed in Section 18.1. We will return to this distinction throughout our analysis.

18.1 Market structure and incentives to innovate

In his classic work *Capitalism, Socialism, and Democracy* (1943), Schumpeter stressed the link between market structure and R&D. His first contention was about the necessity of tolerating the creation of monopolies as a way to encourage the innovation process. This argument is nothing but the economic rationale behind the legal protection of intellectual property that we develop in the next chapter. Schumpeter's second conjecture was that large firms are better equipped to undertake R&D than smaller ones. The best way to support this conjecture is probably to say that large firms have a larger *capacity* to undertake R&D, insofar as they can deal more efficiently with the three market failures observed in innovative markets: externalities, indivisibilities and uncertainty.[a] As far as externalities are concerned, large firms are likely to have fewer competitors able to imitate their innovation. In terms of indivisibilities, large firms are more qualified to exploit increasing returns in R&D. Finally, regarding uncertainty, large firms are more diversified and, hence, more willing to take risks. However, it is not clear whether large firms, because of their monopoly power, also have larger *incentives* to undertake R&D.

[2] See Maurer and Scotchmer (2004, 2006).
[a] We detail these three market failures in Chapter 19.

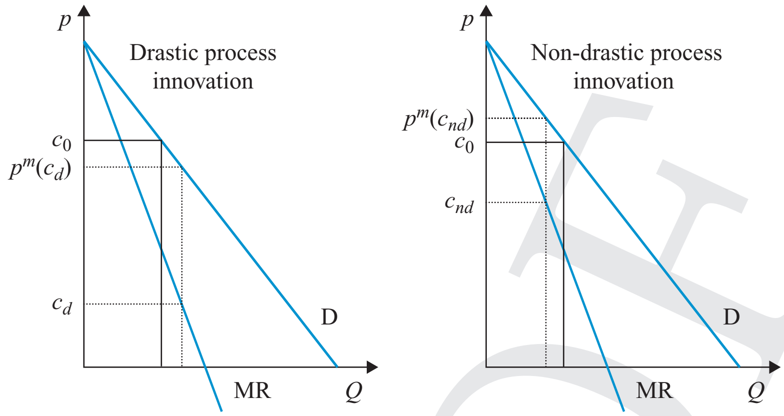

Figure 18.1 Drastic and non-drastic process innovations

To assess the effect of market structure on the incentives for R&D, we proceed in three steps. First, we focus on the *profit incentive* to innovate, leaving aside any strategic consideration. Basically, we want to answer the following question: which market structure, monopoly, oligopoly or perfect competition, provides firms with the highest incentives to undertake R&D? Second, we introduce some *strategic incentives* to innovate by considering either potential competition (between an incumbent and a potential entrant) or actual competition (between oligopolists). Further strategic motivations will be considered in the next two sections.

As a preliminary to our analysis, let us define a number of concepts that will be used throughout Part VII. First, we make a distinction between product and process innovation. A *process* innovation is the generation, introduction and diffusion of a new production process (with the products remaining unchanged). A *product* innovation is the generation, introduction and diffusion of a new product (with the production process remaining unchanged).

We will focus here mainly on process innovations and we will often represent them in our models simply through a reduction of the marginal costs of production for a given product. Why do we leave product innovations aside? Firstly, because we have already addressed product introduction and product positioning in Chapter 5. Secondly, because it can be argued that a product innovation is nothing but an extreme case of a process innovation. That is, the new product already existed but was simply too expensive to produce; so, it took a process innovation to make the new product available.

As for process innovations, it is useful to classify them according to their impact on the market structure. A process innovation is said to be *drastic* (or *major*) if it reduces costs to such an extent that it allows the innovator to behave as a monopolist without being constrained by price competition in the industry. Otherwise, the innovation is said to be *non-drastic* (or *minor*); in this case, the innovator may gain some cost advantage over its rivals but competition constrains the innovator.

Figure 18.1 illustrates the distinction. Consider the market for a homogeneous product. A number of firms produce this product at constant marginal cost c_0 and compete in prices. Suppose that some firm discovers a process innovation that reduces the marginal cost of production. In the left panel of the figure, the innovation reduces the cost from c_0 to c_d. One observes that the cost reduction is so large that the monopoly price corresponding to the new cost c_d, $p^m(c_d)$, falls below c_0. In that case, the innovator can fix the monopoly price without fear of competition from the

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

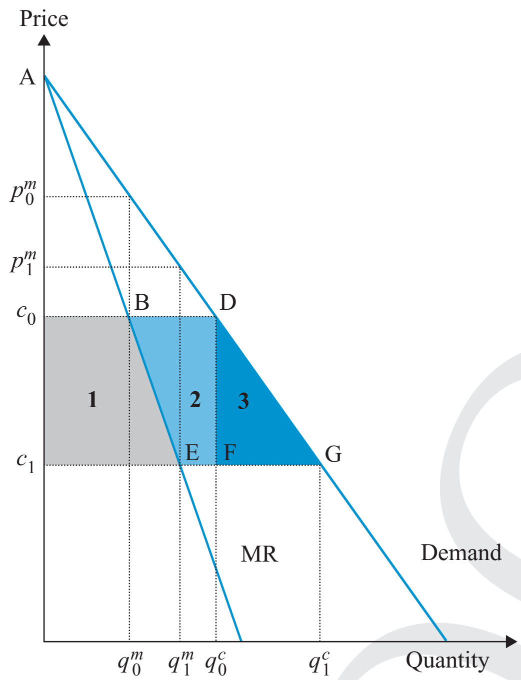

Figure 18.2 Incentives to innovate as a function of market structure

other firms and the innovation is thus drastic. The right panel illustrates the case of a non-drastic innovation: innovation reduces marginal cost only to c_{nd} and because $p^m(c_{nd}) > c_0$, the innovator cannot behave as a monopolist and is constrained by the price competition of the rival firms.

18.1.1 Monopoly versus perfect competition: the replacement effect

To assess the profit incentive to innovate, we can ask the following question, in line with the seminal analysis of Arrow (1962): how much is a firm willing to pay for an innovation that it would be the only one to use? Here, we implicitly refer to a creative environment in which ideas are scarce. For simplicity, consider a non-drastic (minor) process innovation that is protected by a patent of infinite length. This process innovation lowers the constant marginal cost of a particular good from c_0 to $c_1 < c_0$. To assess how much a firm would pay to acquire this innovation (knowing that it will become its sole user), we use the graphic analysis depicted in Figure 18.2.

Consider first an initial *competitive situation*. Prior to the innovation, the quantity q_0^c is sold at price c_0 and all firms earn zero profit. The firm that obtains the new technology produces at cost c_1 and, as the innovation is non-drastic, this firm is constrained to charge $p = c_0 - \varepsilon$ (with ε arbitrarily small) instead of the monopoly price p_1^m (as the other firms are able to charge a price as low as $c_0 < p_1^m$). As the innovator sells the quantity q_0^c, its profit is equal to $\pi^c = q_0^c(c_0 - c_1)$. That is, the per-period value placed by a competitive firm on the innovation is represented in Figure 18.2 by the surface of the rectangle $c_0 DFc_1$, that is, by the sum of areas 1 and 2.[b]

[b] Given that the innovator is awarded a patent of unlimited duration, the incentive to innovate in the competitive situation is given by the total present value of the flow of profits. That is, denoting the interest rate by r, $V^c = \sum_{t=0}^{\infty}(1+r)^{-t}\pi^c = (1/r)(1+r)\pi^c$ (using discrete-time discounting) or $V^c = \int_0^{\infty} e^{-rt}\pi^c\, dt = (1/r)\pi^c$ (using continuous-time discounting).

Consider now an initial *monopoly situation*. We assume that the monopolist faces no threat of entry and is thus the only firm which can benefit from the innovation. Prior to the innovation, the monopolist produces at cost c_0 and its optimum is to sell a quantity q_0^m at price p_0^m. Its profit is equal to $\pi_0^m = q_0^m(p_0^m - c)$. This profit can be measured in Figure 18.2 in two alternative ways. First, it can be measured by the area of the rectangle with base q_0^m and height $(p_0^m - c)$. Equivalently, recalling the principles of integration, profit can be measured by the area below the marginal revenue curve and above the marginal cost curve (as this area corresponds to the difference between total revenues and total costs); that is, the monopolist's profit with cost c_0 is also measured by the area of the triangle Ac_0B. After lowering its cost to c_1, the monopolist finds it optimal to sell a quantity q_1^m at price p_1^m, resulting in a profit equal to $\pi_1^m = q_1^m(p_1^m - c)$. Applying our second way to measure profit, we have that the monopolist's profit after the innovation corresponds to the area of the triangle Ac_1E. We can now gauge the per-period value placed by a monopoly on the innovation as the profit increases, $\pi_1^m - \pi_0^m$. In Figure 18.2, this value is equal to the difference between the triangles Ac_1E and Ac_0B, which is given by area 1 (i.e., area c_1c_0BE). Recalling that a competitive firm is willing to pay up to the areas 1 *and* 2, we reach the following conclusion.

Lesson 18.1 **A competitive firm places a larger value on a minor process innovation than a monopoly does.**

The intuition behind this result is quite straightforward and comes from the comparison of the pre-innovation situations: prior to the innovation, the monopolist already earns a positive profit, whereas the competitive firm just recoups its costs. This is known as the *replacement effect*: for the competitive firm, the innovation creates a brand new profit opportunity but for the monopolist, the innovation just 'replaces' an existing profit by a larger one. Innovation entails thus a larger opportunity cost for the monopolist than for a competitive firm. Note that the same argument applies even more clearly for the case of a drastic innovation. Here, the (per-period) incentives to innovate for a competitive firm and for a monopolist are respectively given by π_1^m and $\pi_1^m - \pi_0^m$.

As illustrated in Case 18.1, the replacement effect can also explain why a firm that is active on several markets tends to direct its R&D investments more towards the markets on which it faces more competition.

Case 18.1 Microsoft's incentives to innovate

Arrow's argument about the replacement effect can be extended to a multiproduct firm. Following the argument, a multiproduct firm would have higher incentives to innovate on the market segments where it faces competition than on those segments where it enjoys significant market power. The following quote, taken from an analysis of Microsoft's launch of the Xbox in 2005, perfectly illustrates the previous conjecture. 'It is surely no coincidence that Microsoft's hidden ability to innovate has become apparent only in a market in which it is the underdog and faces fierce competition. Microsoft is far less innovative in its core businesses, in which it has a monopoly (in Windows) and a near monopoly (in Office). But in the new

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

markets of gaming, mobile devices and television set-top boxes, Microsoft has been unable to exploit its Windows monopoly other than indirectly – it has financed the company's expensive forays into pasture new.'[3]

It is worth noting that the value the competitive firm places on the innovation still falls short of the social value of the innovation. Indeed, a benevolent social planner would be willing to pay up to the increase in social surplus that the innovation generates. The per-period social value of the innovation is represented in Figure 18.2 by the difference between the areas of triangle c_1AG (social surplus under cost c_1) and triangle c_0AD (social surplus under cost c_0), that is, by the sum of areas 1, 2 and 3. The reason is that the competitive firm fails to appropriate the increase in consumer surplus (i.e., area 3).

18.1.2 Incentives to innovate in oligopolies

As oligopolies are intermediate market structures between the two extremes of monopoly and perfect competition, one could be tempted to infer from the previous argument that incentives to innovate in oligopolies stand somewhere between the low incentives of the monopolist and the high incentives of the perfectly competitive firm. However, this conjecture turns out to be wrong. To show this, we reconsider our initial question (i.e., how much is a firm willing to pay for an innovation that it would be the only one to use?) and examine how its answer is affected by the intensity of competition.

There is no obvious way to address this issue as the intensity of competition can be measured in different ways: either by the number of firms on the market, or by the degree of product substitutability, or by the nature of competition (as we showed in Chapter 3, price competition leads to more competitive outcomes than quantity competition). Here, we take the number of firms as a measure of the intensity of competition. Using the simple linear Cournot model with n firms introduced in Chapter 3, we want to show that *the profit incentive to innovate may follow an inverse U-shape as the number of firms in the industry increases.* As a consequence, Cournot competition with an adequate number of firms may lead to larger incentives to innovate than both perfect competition and monopoly.[4]

Suppose that n firms compete à la Cournot on the market of a homogeneous product. Suppose further that both the demand and the cost functions are linear. Specifically, we take $P(q) = a - q$ (with $a > 0$ and $q = \sum_i q_i$) and $C_i(q_i) = c_i q_i$ (with $0 \leq c_i < a$ $\forall i = 1, \ldots, n$). In Chapter 3, we showed that firm i's equilibrium profits at the Cournot equilibrium are equal to

$$\pi_i^* = \left(\frac{a - nc_i + \sum_{j \neq i} c_j}{n+1} \right)^2 . \tag{18.1}$$

Initially, all firms produce at cost $c_i = c_0$. An independent research lab finds a process innovation that reduces the constant marginal cost of production from c_0 to $c_1 < c_0$. We assume

[3] Taken from 'The meaning of XBox', *The Economist*, 24 November 2005.

[4] A number of papers examine the link between various measures of competition and the incentive to innovate. See, for example, Delbono and Denicolo (1990), Bester and Petrakis (1993) and Qiu (1997). For a comprehensive analysis within a unified framework, see Belleflamme and Vergari (2011).

that the innovation is non-drastic. That is, the monopoly price corresponding to c_1 is larger than the initial cost c_0: $(a + c_1)/2 > c_0$, which is equivalent to $c_0 - c_1 < a - c_0$. In the latter inequality, $c_0 - c_1$ measures the absolute size of the innovation (i.e., the cost reduction) and $a - c_0$ measures the size of the initial market (i.e., the difference between the largest price consumers are willing to pay, a, and the pre-innovation marginal cost). We can thus measure the *relative size of the innovation* by the ratio

$$\psi \equiv \frac{c_0 - c_1}{a - c_0} \quad \text{with} \quad 0 < \psi < 1.$$

The pre-innovation profit, π_{pre}, is easily computed by setting $c_i = c_j = c_0$ in expression (18.1). As for the post-innovation situation, we are only interested in the profit accruing to the innovator, which we denote π_{post}; we also compute it easily by setting $c_i = c_1$ and $c_j = c_0$ $(\forall j \neq i)$ in expression (18.1). Accordingly, we have

$$\pi_{pre} = \left(\frac{a - c_0}{n+1}\right)^2 \quad \text{and} \quad \pi_{post} = \left(\frac{a - nc_1 + (n-1)\,c_0}{n+1}\right)^2.$$

We measure the profit incentive to innovate as the extra profit that accrues to the innovator with respect to the pre-innovation situation: $PI \equiv \pi_{post} - \pi_{pre}$. After some manipulations and emphasizing that the profit incentive depends on the number of firms, we have

$$PI(n) = \tfrac{n}{(n+1)^2}(2 + n\psi)\psi(a - c_0)^2.$$

We want now to assess how the profit incentive to innovate changes as the number of firms in the industry increases. To this end, we compute

$$\begin{aligned} PI(n+1) - PI(n) &= \left[\tfrac{n+1}{(n+2)^2}\left(2 + (n+1)\,\psi\right) - \tfrac{n}{(n+1)^2}\left(2 + n\psi\right)\right]\psi\,(a - c_0)^2 \\ &= \frac{\left(2n^2 + 4n + 1\right)\psi - 2\left(n^2 + n - 1\right)}{(n+2)^2\,(n+1)^2}\,\psi\,(a - c_0)^2. \end{aligned}$$

We understand from the latter expression that *an increase in the number of firms raises the profit incentive to innovate if the relative size of the innovation is large enough*. Formally, defining

$$\hat{\psi}(n) \equiv \frac{2(n^2 + n - 1)}{2n^2 + 4n + 1},$$

we have that $PI(n+1) > PI(n)$ if and only if $\psi > \hat{\psi}\,(n)$. We note that $\hat{\psi}\,(n)$ increases with n and tends to 1 as n tends to infinity. Therefore, there exists a finite size of the Cournot industry that maximizes the profit incentive to innovate. In particular, monopoly leads to the largest incentive to innovate if $PI\,(2) < PI\,(1)$ or $\psi < \hat{\psi}\,(1) = 2/7 \simeq 0.286$; duopoly leads to the largest incentive to innovate if $\hat{\psi}\,(1) \simeq 0.286 < \psi < \hat{\psi}\,(3) = 10/17 \simeq 0.588$; and so on and so forth.

Lesson 18.2 **In a Cournot industry with a homogeneous product, the market structure that gives the largest profit incentive to innovate is monopoly when the innovation size is not too large; it is oligopoly otherwise (and the 'ideal' number of firms in the industry increases with the innovation size).**

To understand the previous result, note that an increase in the number of firms in the industry has two opposite effects on the profit incentive to innovate. On the one hand, there is a competition effect: a larger number of firms reduces profits both for the innovating and for the non-innovating firms. On the other hand, there is a competitive advantage: the larger n, the larger the number of rival firms producing in a less efficient way. The coexistence of these two conflicting forces explains why the incentive to innovate first increases and then decreases as the industry gradually expands.

18.1.3 Patent licensing

In the next chapter, we will explain how patents (and more generally intellectual property rights) act as an incentive mechanism to stimulate innovative activities: an innovator who obtains a patent for his innovation is granted exclusive rights over this innovation, which allows him to secure a flow of profits and, thereby, to possibly recoup the cost of the initial R&D investment. What we want to stress here is that the exclusive rights granted by a patent are also *transferable rights* (they can be transferred, licensed, rented or mortgaged to third parties). This transferability is economically very important as it aims to ensure that innovations and artistic creations are used by the agents who value them most. Transferability is also important because it gives an additional source of profit to the innovator: the innovator can earn profit through his own working of the patent but also by licensing the patent. The payment structure of patent licences can take the following forms: (i) royalties (which are computed either per unit of output produced with the patented technology or as a quota of the licensee's revenues – so-called 'ad valorem' royalties); (ii) a fixed fee that is independent of the quantity produced with the patented technology; or (iii) a combination of the previous two options. The patent holder (the 'patentee' or 'licensor') can choose among these three options and can also decide either to allow any firm to purchase a licence or to auction a limited number of licences.

From a social viewpoint, licences have the clear positive impact of increasing the diffusion and use of knowledge. However, one needs to evaluate the impact of licences on the incentives to innovate. We examine this issue in two different contexts according to whether the patentee aims to license its innovation outside or inside his own industry.

18.1.4 Licensing by an outside innovator

When the patentee licenses his innovation outside his own industry, he will not compete with his potential licensees on the product market. One can think of the innovator as, for example, a research lab that is not able (or not willing) to exploit the innovation itself and is thus transferring its rights to manufacturing firms. Note that this is the scenario that we have analysed so far, although in a partial way: we looked for the maximum fixed fee that a firm would be willing to

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

pay to become the exclusive licensee. We now want to extend this analysis by letting the innovator grant multiple licences and choose between fixed fees and per-unit royalties.[5]

Consider a process innovation that reduces the constant marginal cost of production of some final product from c to $c - x$, with $0 < x < c$. There are n firms (indexed by i) producing the final product and the market demand is linear: $P(q) = a - q$, with $a > 0$ and $q = \sum_i q_i$. Recall our distinction between drastic (or major) and non-drastic (or minor) innovations: an innovation is drastic (resp. non-drastic) if the post-innovation monopoly price is below (resp. above) the pre-innovation competitive price. In the present case, the post-innovation monopoly price is given by $(a + c - x)/2$, while the pre-innovation competitive price is c. Comparing the two, we find that the innovation is drastic if $x > a - c$ and non-drastic if $x < a - c$. Given that $x < c$, we impose $c > a/2$ to make the case of drastic innovations possible.

If the n manufacturing firms compete à la Bertrand, the problem of the outside innovator is rather simple. First, there is no point in granting more than one licence: for a firm to make a positive profit at the Bertrand equilibrium, it has to be the only firm with lower production cost. If the innovation is drastic, an exclusive licence is worth the monopoly profit corresponding to cost $c - x$, that is, $\pi^m = (a - c + x)^2/4$. The outside innovator can capture this value by setting a fixed fee $F = \pi^m$ or by setting a royalty equal to the monopoly margin, $r = (a + c - x)/2 - (c - x) = (a - c + x)/2$. If the innovation is non-drastic, the licensee would set its price just below c and sell a quantity $q \simeq a - c$, thereby securing a profit almost equal to $\bar{\pi} = x(a - c)$. The innovator can capture this profit equivalently through a fixed fee $F = \bar{\pi}$ or through a royalty $r = x$. We thus conclude the following.

Lesson 18.3 **If an outside innovator licenses his process innovation to an industry competing à la Bertrand, he chooses to grant one licence and is indifferent between a fixed fee and a royalty.**

Suppose now that the manufacturing firms compete à la Cournot. To simplify our analysis, we restrict the attention to a duopoly.[6] We analyse the following three-stage game. In the first stage, the innovator sets a fixed licensing fee or a per-unit output royalty rate; in the second stage, the two manufacturing firms decide simultaneously whether to accept the offer or not; in the third stage, the duopolists compete à la Cournot on the product market. Our goal is to characterize the subgame-perfect equilibria of this game.

Assume first that the innovator chooses fixed-fee licensing at the first stage. At the third stage, the equilibrium of the linear Cournot duopoly with given costs (c_i, c_j) yields the following quantity and profit for firm i (we refer the reader to Chapter 3 for the details):

$$q_i^*(c_i, c_j) = \tfrac{1}{3}(a - 2c_i + c_j) \text{ and } \pi_i^*(c_i, c_j) = \tfrac{1}{9}(a - 2c_i + c_j)^2.$$

[5] Although ad valorem royalties seem to be more widespread in practice than per-unit royalties, relatively little attention has been devoted to them in the economic literature. Bousquet *et al*. (1998) argue that uncertainty on the demand side may lead an outside innovator to prefer ad valorem royalties to per-unit royalties. For a review of the recent literature on ad valorem royalties, see Colombo and Filippini (2014).

[6] We follow here Kamien and Taumann (1986), who show that the results can be generalized to an n-firm symmetric Cournot oligopoly.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

These results are valid as long as the high-cost firm finds it profitable to produce a positive quantity. Suppose that $c_j > c_i$; then $q_j^*(c_i, c_j) \geq 0$ as long as $c_j \leq (a + c_i)/2$. If firm i is a licensee and firm j is not (meaning that $c_i = c - x$ and $c_j = c$), then the latter condition boils down to $x \leq a - c$. That is, if the patentee grants only one licence, both firms are active at the third stage if the innovation is non-drastic. If the innovation is drastic, then the licensee is a de facto monopolist, achieving a profit of $\pi^m(c - x) = (a - c + x)^2/4$.

At the second stage, we can compute the firms' willingness to pay to be an exclusive licensee (denoted F_1) or to be one of two licensees (denoted F_2) as:

$$F_1 = \begin{cases} \pi_i^*(c - x, c) - \pi_i^*(c, c) & \text{(non-drastic innovation),} \\ \pi^m(c - x) - \pi_i^*(c, c) & \text{(drastic innovation),} \end{cases}$$

$$F_2 = \begin{cases} \pi_i^*(c - x, c - x) - \pi_i^*(c, c - x) & \text{(non-drastic innovation),} \\ \pi_i^*(c - x, c - x) - 0 & \text{(drastic innovation).} \end{cases}$$

The innovator will make a take-it-or-leave-it offer to the firms and set a fixed fee up to their willingness to pay. It is then optimal to grant an exclusive licence if $F_1 > 2F_2$. Using the above expressions, one can compute that for drastic and non-drastic innovations, the difference $(F_1 - 2F_2)$ has the same sign as the difference $x - (a - c)$. That is, it is more profitable to grant an exclusive licence if the innovation is drastic, but to grant two licences if it is non-drastic.

Let us now repeat the analysis in the case where the innovator chooses royalty licensing at the first stage. Let r denote the amount of royalty that licensees have to pay per unit of output; the marginal cost of production for a licensee is then equal to $c - x + r$. As a result, at the second stage, both firms accept the licensing contract if $r \leq x$, but none of them does if $r > x$. If both firms accept, the total equilibrium quantity at stage 3 is $2q_i^*(c - x + r, c - x + r) = \frac{2}{3}(a - c + x - r)$. In stage 1, the innovator's maximization problem is thus

$$\max_r r\frac{2}{3}(a - c + x - r) \text{ subject to } r \leq x.$$

The unconstrained maximum is $r^* = (a - c + x)/2$; we have that $r^* \leq x$ if and only if $x \geq a - c$. Hence, if the innovation is drastic ($x \geq a - c$), the constraint is satisfied and the innovator sets $r = r^*$; the corresponding licensing revenues are computed as $R_d = \frac{1}{6}(a - c + x)^2$ (with d for drastic). Otherwise, if the innovation is non-drastic ($x < a - c$), we have a corner solution: the innovator sets $r = x$, yielding licensing revenues equal to $R_n = \frac{2}{3}x(a - c)$ (with n for non-drastic).

We are now in a position to solve for the innovator's optimal behaviour at the first stage of the game. For drastic innovations, the comparison is between $F_1 = \pi^m(c - x) - \pi_i^*(c, c)$ and R_d; a fixed-fee contract is more profitable than a royalty-based contract if $\frac{1}{4}(a - c + x)^2 - \frac{1}{9}(a - c)^2 > \frac{1}{6}(a - c + x)^2$, which is equivalent to

$$3x^2 + 6(a - c)x - (a - c)^2 > 0.$$

The latter inequality is clearly satisfied as the left-hand side increases with x and $x \geq a - c$ (drastic innovation). For non-drastic innovations, the innovator compares $2F_2 = 2(\pi_i^*(c - x, c - x) - \pi_i^*(c, c - x))$ to R_n and prefers a fixed fee if

$$2F_2 > R_n \Leftrightarrow \frac{8}{9}x(a - c) > \frac{2}{3}x(a - c),$$

which is always true. As the previous results can be shown to hold if the manufacturing industry is made up of an arbitrary number of symmetric firms, we can conclude the following.

Lesson 18.4 **If an outside innovator licenses his process innovation to an industry competing à la Cournot, he always prefers fixed-fee licensing to royalty licensing. He licenses drastic innovations to only one firm and non-drastic innovations to more than one firm.**

The intuition for the dominance of fixed-fee licensing is straightforward: licensees are more efficient under fixed-fee licensing than under royalty licensing (as royalties are added to their marginal cost of production); hence, more profit is generated, which the patentee can reap via the fixed fee.

However, this theoretical prediction does not square with the empirical observations according to which the two forms of licensing are equally prevalent. A first way to reconcile theory and facts is to twist the previous game a little by assuming that at the first stage, the patentee not only announces a royalty or a fixed fee but also decides how many licensees he will accept. Under this assumption, when the innovation is drastic, the profit-maximizing royalty scheme consists of preventing one of the two firms from getting a licence; as the exclusive licensee will monopolize the industry, licensing revenues are equal to $R_d' = \frac{1}{4}(a-c+x)^2 > R_d = \frac{1}{6}(a-c+x)^2$; it is then immediate to check that $R_d' > F_1$, implying that a per-unit royalty dominates a fixed fee for drastic innovations. A second way to show that royalties may be preferred is to introduce cost asymmetries into the previous model. Suppose for instance that (before any licensing) firm 1 is more efficient than firm 2 ($0 < c_1 < c_2 < a$), and that obtaining a licence allows firm i to lower its cost from c_i to $c_i - x$ (with $x < c_1$). In this setting, royalty licensing may dominate fixed-fee licensing. For this to happen, the difference in pre-licensing efficiency (i.e., $c_2 - c_1$) must be large enough. The reason is the following: when c_2 is sufficiently larger than c_1, and when the innovator wants to license to both firms, the fee needed to convince firm 2 to accept the licence is so low that royalty licensing becomes preferable.[7]

18.1.5 Licensing by an inside innovator

We consider now the case where the innovator belongs to the manufacturing industry. The question we address then is: does an 'inside innovator' have an incentive to license its patented discovery to some or all of its rivals? As we now show, the answer depends on the innovation size and on the market structure. Consider first the case of *drastic innovations*. Because such innovations allow the patentee to force its competitors out of the market and become a monopolist, it is intuitively clear that the patentee has no incentive to grant licences, as this will reintroduce competition in the market and, thereby, lower its profits. This conclusion holds whether competition on the market is Cournot or Bertrand.[c]

[7] This result is due to Wang and Yang (2004). Evidence about the prevalence of royalty and fixed-fee licensing can be found, for example, in Jensen and Thursby (2001).

[c] Even if the patentee is able to extract almost the entire licensees' profits via a fixed fee or a royalty contract, the sum of oligopolists' profits is less than the monopoly profits when products are not differentiated. By contrast, under sufficient product differentiation, the patentee has an incentive to license.

Let us now turn to the more interesting case of *non-drastic innovations*. As above, we assume that the inverse demand is given by $p = a - Q$, and that the innovation has the effect of reducing the constant marginal cost from c (with $c < a$) to $c - x$ (with $0 < x < a - c$), so that the innovation is non-drastic; there are n symmetric firms in the industry and we identify the innovator as firm 1. Consider first Bertrand competition. In this case, it is easy to see that the innovator does not gain from licensing. When the innovator does not license the innovation, we saw above that its profit-maximizing action is to set a price just below the other firms' cost ($p = c - \varepsilon$), thereby serving the whole market and securing a margin of $(x - \varepsilon)$ on each unit sold. Alternatively, the innovator could license the innovation to any rival via a royalty contract. In that case, the potentially profit-maximizing royalty rate, r, is such that the rivals' new cost, $c - x + r$, is just below their original cost c: that is, $r = x - \varepsilon$. Then, the patentee and its licensees share the market at price $p = c - \varepsilon$; the quantity sold is the same as in the absence of licensing; the innovator secures a margin of $(x - \varepsilon)$ on the units it sells and collects a royalty $r = x - \varepsilon$ on the units sold by the licensees. In total, the innovator achieves thus exactly the same profit as in the absence of licensing, which proves our claim. A similar reasoning holds for a fixed-fee contract. Here, the best the innovator could do would be to sell only one licence (as two licensees would make zero profits) and forego the right to use the innovation himself; the licensee would then serve the whole market, taking the place of the innovator. In this extreme situation, the innovator would collect the monopoly profit from the licensee instead of securing it by its own production; in the presence of transaction costs – even negligible – this operation would not be profitable for the patentee.

Lesson 18.5 **It is not profitable for an inside innovator to license its process innovation to its (symmetric) industry rivals (i) when the innovation is drastic or (ii) when the innovation is non-drastic and Bertrand competition prevails on the product market.**

Consider next Cournot competition. Let us first show that contrary to the case of an outside innovator, royalty licensing is now the preferred option. Recall that the outside innovator prefers fixed-fee licensing because it makes the licensees more efficient, thereby allowing them to generate more profits, which can later be captured through the fixed fee. Yet, the very same reason makes fixed-fee licensing *less* advantageous for an inside innovator as he is in direct competition with the licensees. Royalty licensing, in contrast, presents the double advantage of securing licensing revenues without damaging the innovator's competitive advantage.

We establish this result more formally in the duopoly case. If the innovator does not license the innovation, his profit is $\pi_{1n} \equiv \pi_i^*(c - x, c)$. If he licenses the innovation to his rival via a fixed fee, the highest fee he can charge is $F_2 = \pi_i^*(c - x, c - x) - \pi_i^*(c, c - x)$; yet, as the licensee can now produce at cost $c - x$, the innovator's profit from production is reduced to $\pi_i^*(c - x, c - x)$. It follows that the innovator's total profit (including licensing revenue) is equal to $\pi_{1f} \equiv 2\pi_i^*(c - x, c - x) - \pi_i^*(c, c - x)$. Finally, under royalty licensing, the innovator charges a royalty $r = x$, keeps his profit from production unchanged and obtains royalty revenues equal to $xq_i^*(c, c - x)$; hence, his total profit is $\pi_{1r} \equiv \pi_i^*(c - x, c) + xq_i^*(c, c - x)$.

A first obvious conclusion is that royalty licensing is always more profitable than granting no licence as it keeps the profit from production unchanged and adds royalty payments:

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

$\pi_{1r} = \pi_{1n} + xq_i^*(c, c-x)$. Second, we also see that fixed-fee licensing may be less profitable than granting no licence; this is so if $\pi_{1f} < \pi_{1n}$, which can be rewritten as $F_2 < \pi_i^*(c-x, c) - \pi_i^*(c-x, c-x)$. In words, fixed-fee licensing makes the innovator worse off if what the licensee gains with the licence (i.e., F_2) is less than what the innovator loses because of increased competition. In our example, the latter condition is equivalent to $x > \frac{2}{3}(a-c)$, that is, if the innovation is large enough. Now, even if fixed-fee licensing is profitable, it is easy to show that royalty licensing is always more profitable; we indeed compute that $\pi_{1r} - \pi_{1f} = \frac{1}{9}x(a-c) > 0$.

In light of this finding, we focus now on royalty licensing and address the question of the number of licences that the innovator will find it optimal to grant.[8] As above, we analyse the following three-stage game: in the first stage, the incumbent innovator selects a royalty rate r at which it will license its new technology; in the second stage, the other firms decide simultaneously whether or not to become licensees; in the third stage, the firms in the industry (the innovator, the licensed and unlicensed firms) engage in Cournot competition.

If the innovator decides to license its technology, it will select a royalty no larger than the cost reduction the innovation entails: $r \leq x$. At such a rate, every firm at stage 2 is willing to become a licensee since this weakly reduces its marginal cost from c to $c - x + r$. Therefore, at stage 3, competition takes place among one firm with cost $c - x$ and $n - 1$ firms with cost $c - x + r$. We showed in Chapter 3 that the Cournot equilibrium quantity and profit for a typical firm k under this setting are respectively given by

$$q_k^* = \frac{1}{n+1}\left(a - nc_k + \sum_{j \neq k} c_j\right), \text{ and } \pi_k^* = \left(q_k^*\right)^2. \tag{18.2}$$

As for the innovator, its marginal cost is $c - x$ and the sum of the rivals' marginal costs is equal to $(n-1)(c-x+r)$; as for a typical licensee, its marginal cost is $c - x + r$ and the sum of the rivals' marginal costs is equal to $(n-2)(c-x+r) + (c-x)$. Substituting these values into Equation (18.2), we can derive the equilibrium quantity for the innovator and for the licensees respectively as

$$q_1^* = \frac{a - c + x + r(n-1)}{n+1} \quad \text{and} \quad q_{lic}^* = \frac{a - c + x - 2r}{n+1},$$

where $q_{lic}^* > 0$ as $r < x < a - c$. The innovator's total profit is then computed as

$$\pi_{1r} = \left(q_1^*\right)^2 + r(n-1)q_{lic}^*.$$

Let us now compute the optimal royalty rate level. Deriving the innovator's profit with respect to r, we get

$$\frac{\partial \pi_{1r}}{\partial r} = \frac{(n-1)(n+3)}{(n+1)^2}(a - c + x - 2r) > 0,$$

which implies that $r^* = x$. Substituting $r = x$ in the expression for π_{1r}, we obtain

$$\pi_{1r}^* = \frac{(a - c + nx)^2}{(n+1)^2} + (n-1)x\frac{(a-c-x)}{n+1},$$

[8] We follow the analysis of Kamien and Tauman (2002). The authors also consider two alternative modes of patent licensing: the auction method and the fixed-fee method. They show that if the size of the innovation is large enough, precisely for $x > (a-c)/(n+2)$, the inside innovator prefers to license its innovation by means of a royalty rather than an auction or a fixed fee.

where the first term is the profit from production, π_{1n}, that the innovator would achieve if he was not granting any licence. The latter situation would indeed involve competition among one firm with cost $c - x$ and $(n - 1)$ firms with cost c. We therefore see immediately that the inside innovator gains from licensing the innovation to all his competitors. The fact that $\pi^*_{1r} - \pi_{1n} > 0$ is easy to understand. Whether the innovator licenses the innovation at $r = x$ or whether he keeps the innovation for his exclusive use does not affect the competitive situation: in both cases, the innovator has marginal cost $c - x$ and his rivals have marginal cost c. So, the innovator makes the same profit from production in both situations (i.e., π_{1n}). However, when he licenses the innovation, the innovator also collects royalties, which strictly improves his total profits.

Note that society also gains from licensing. Indeed, with $r = x$, total ouput is the same with or without licensing, so that consumers are as well off; the other $(n - 1)$ firms have the same profit but the innovator strictly increases its profit. Total surplus is thus strictly larger.

Lesson 18.6 **In the case of quantity competition on the product market, it is always profitable for an inside innovator to license a non-drastic process innovation to all its industry rivals through a royalty contract. Licensing also increases total surplus.**

18.2 When innovation affects market structure

The previous section dealt with creative environments where ideas are scarce. It was thus appropriate to represent the innovation process as external to the industry under review and to measure how much a given firm in the industry is willing to pay for an innovation. In contrast, in creative environments in which ideas are common knowledge, several firms have the simultaneous opportunity to achieve competing innovations. Innovation therefore becomes a competitive tool in itself and its effects on the market structure need to be assessed.

To this end, we reconsider incentives to innovate in a monopoly when the threat of entry is taken into account. We want to identify who, between the incumbent firm and the entrant, is the most likely to innovate. The answer to this question will indeed indicate how innovation affects market structure: simply put, if the incumbent innovates, monopoly persists but if the entrant innovates, entry takes place and the industry becomes a duopoly. We address this issue first in a simple static setting where the innovation process is certain (Subsection 18.2.1) and next in a more realistic dynamic setting where the innovation process is uncertain (Subsection 18.2.2). Finally, in Subsection 18.2.3, we consider dynamic R&D competition between two symmetric firms and compare their R&D decisions with the choice of a social planner; the concern here is that R&D competition might give firms too large an incentive to invest in R&D from a social point of view.

18.2.1 Monopoly threatened by entry: the efficiency effect

We consider here that two firms can acquire a given innovation: an incumbent monopoly as well as a potential entrant. If both firms are active in the market, they are assumed to make duopoly profit π^d. One can imagine that the innovation has been discovered by a third firm which cannot

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

(or prefers not to) exploit it itself and auctions it to the highest bidder. The question is thus: who, between the incumbent and the entrant, will place the highest bid for the innovation?

Clearly, each firm is willing to bid up to the difference between the profit it would make if it gets the innovation and the profit it would make if it does not. For the potential entrant, the situation is quite similar to what we had in the previous section: if the potential entrant does not obtain the innovation, it stays out of the market and earns zero profit; on the contrary, with the innovation, the entrant becomes a duopolist with a cost advantage; we denote its profit by $\pi^d(c_1, c_0)$. Hence, the value of the innovation for the entrant is its discounted duopoly profit $V_E = (1/r)\,\pi^d(c_1, c_0)$, where r is the continuous-time discount factor.

Things are a bit more complicated for the incumbent as it has to take into account what happens if the entrant is the one who successfully innovates. In that case, the incumbent would become a duopolist with a cost disadvantage and would earn a profit equal to $\pi^d(c_0, c_1)$. However, if the incumbent successfully innovates, it avoids entry and earns the monopoly profit $\pi^m(c_1)$. As a result, the incumbent places the following value on the innovation: $V_I = (1/r)[\pi^m(c_1) - \pi^d(c_0, c_1)]$.

Comparing the two values, we see that the incumbent has a higher incentive to acquire the innovation if the following condition is met:

$$V_I > V_E \Leftrightarrow \pi^m(c_1) > \pi^d(c_0, c_1) + \pi^d(c_1, c_0). \tag{18.3}$$

This condition is satisfied when the products sold by the two firms are close substitutes. In that case, a monopoly with a low cost earns a higher profit than two non-colluding duopolists can earn together, especially if one of them produces at a higher unit cost. This is so because the monopolist is always able to mimic the choices made by the duopolists; so, if it chooses otherwise, it is because such choices can be improved upon in terms of profit. However, the previous argument might no longer hold if the entrant is bringing a significantly differentiated product to the market. If consumers enjoy variety, the entry of a differentiated product might increase the size of demand in such a way that the inequality in (18.3) is reversed. This is also explained by the fact that the incumbent has less to lose from entry when the entrant's product is not a close substitute (in the limit, if the two products are completely differentiated, $\pi^d(c_0, c_1) = \pi^m(c_0)$ and $\pi^d(c_1, c_0) = \pi^m(c_1)$, and the replacement effect is the only effect at play).

Lesson 18.7 **A monopoly threatened by entry is willing to pay more for a minor innovation than a potential entrant who can produce a close substitute to the monopolist's product.**

The fear of losing its monopoly position provides the incumbent with a stronger incentive, a property known as the *efficiency effect*.[9] It follows that in industries where entry represents an important threat, incumbent firms use innovation as a tool to prolong their monopoly situation over time. As a result, innovation affects market structure, as illustrated by the following two real-life applications.

[9] This was stressed by Gilbert and Newbery (1982).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

Blocking patents

It is common in many industries for the same firm to patent a large number of related processes or products but only use a small number of them. For instance, according to an article published in the *Financial Times* in July 2001, IBM, Philips and Siemens were reported to use only about 40% of their portfolio of patents. Several reasons can be proposed to explain why a number of new processes or products remain under- or unexploited on firms' shelves: duplication of R&D efforts, lack of complementary assets to bring the innovation to market, or poor fit between the innovation and the firm's objectives; such patents are called 'sleeping' (or 'dormant'). Strategy considerations might be invoked as well: unused patents may allow dominant firms to block entry into their market; one talks then of 'blocking' patents. Holding several patents all related to the same process or product would thus create a buffer of protection around the truly valuable patent. It appears then, invoking the efficiency effect argument, that an incumbent firm is willing to pay more to prevent a rival from using an innovation than what this rival is willing to pay to actually use this innovation.

Pay-for-delay agreements

'Pay-for-delay' agreements are commonplace in the pharmaceutical industry; they involve branded drug makers paying generic groups to delay the launch of lower-cost versions of their drugs. Besides the potential anticompetitive nature of such deals (which is widely discussed in the USA and in Europe), there is the question of the economic rationale of such deals for the involved parties. It does not take too much thinking to understand why the originator pharmaceutical firms propose such deals, as they allow them to prolong their monopoly position. What is more intriguing, at first glance, is that generic producers accept these deals. One can indeed wonder whether it would not be more profitable for them to reap the profits of an earlier launch of their generic drugs rather than accepting the money that is offered to them if they delay. We can use the above analysis to shed some light on this issue. The maximum amount that the originator is willing to pay to delay entry is the difference between the monopoly and the duopoly profit: $\pi^m - \pi^d$. In contrast, the generic producer demands a payment at least equal to the duopoly profit π^d to accept not to enter. An agreement can be found between the two parties if $\pi^m - \pi^d > \pi^d$ or $\pi^m > 2\pi^d$, which is satisfied in many markets as we argued above.[d]

18.2.2 Asymmetric patent races: replacement and efficiency effects

The previous analysis abstracted away two important dimensions of R&D competition, namely *time* and *uncertainty*. Indeed, it was implicitly assumed that an investment in R&D translated immediately, and with certainty, into some process innovation. In many situations, however, the innovation process is uncertain and the real objective of R&D competition is to be the first to come up with an innovation (so as to outperform the rival firms). R&D competition takes thus the form of a *race for a patent*. In such a race, firms have to decide not only about the intensity of their R&D investment (so as to increase their chances of finding the innovation) but also about the timing of their R&D investments (so as to be the first to innovate).

[d] Any standard homogeneous-product oligopoly model satisfies this property.

To enrich our analysis of the effect of R&D competition on market structure, we examine here an asymmetric patent race between an incumbent firm and an entrant.[10] Recall the elements we have gathered so far. On the one hand, we invoked the *replacement effect* to argue that monopoly power acts as a strong disincentive to innovate. On the other hand, we predicted the persistence of monopoly on the basis of the *efficiency effect*: in a static setting, a monopolist threatened by entry is willing to pay more for an innovation than a potential entrant. Our goal is now to study the interplay between these two effects in the framework of a patent race. Therefore, we use the same setup as in the previous subsection. Recall that the innovation allows firms to reduce the marginal cost of production from c_0 to $c_1 < c_0$, and that profits (per unit of time; we assume continuous time) are given as in the following table.

		After innovation	
	Before innovation	Incumbent obtains patent	Entrant obtains patent
Incumbent	$\pi^m(c_0) \equiv \pi_0^m$	$\pi^m(c_1) \equiv \pi_1^m$	$\pi^d(c_0, c_1) \equiv \pi_{01}^d$
Entrant	0	0	$\pi^d(c_1, c_0) \equiv \pi_{10}^d$

At each point in time, firm k spends some amount x_k on R&D. We assume a 'memoryless' stochastic process of the Poisson type. This means that a firm's probability of success depends only on its current R&D at any point in time and not on the accumulated stock of investment in R&D. This assumption allows us to abstract away issues related to dynamic investment problems and diffusion of knowledge. The probability of success is formalized as a (twice continuously differentiable) hazard function $h(x)$, which is strictly concave and satisfies some boundary condition, so that we only need to consider interior solutions. To be precise, (i) $h(0) = 0$, (ii) $h' > 0$, (iii) $h'' < 0$, (iv) $\lim_{x\to\infty} h'(x) = 0$ and (v) $\lim_{x\to 0} h'(x) = \infty$. A strategy for firm k consists of specifying its research expenditure for each point in time. Hence, it sets a function $x_k(\cdot)$. Since h is independent of the competitor's history and the competitor's history is payoff-irrelevant, x_k may depend on time only. Yet, since the stochastic process is memoryless, time has no effect. Indeed, at each date t, if neither firm has made a discovery, the game starting at this moment is identical to the initial game. It follows that the equilibrium strategies are constant over time.

We start by deriving the probability that no firm has made a discovery before some date t. Assuming that the patent race starts at date 0, we can use the Poisson process to determine the probability that firm k has been successful by date t.[e] Denoting $\tau(x_k)$ the date of success given R&D expenditure x_k, we have that $\text{Prob}\{\tau(x_k) \leq t\} = 1 - e^{-h(x_k)t}$. Therefore, the probability that at date t neither of the two firms succeeded is equal to $e^{-(h(x_I)+h(x_E))t}$ (where the subscripts I and E refer, respectively, to the incumbent and to the entrant). We can now compute the present discounted value of the expected profit over time for the two firms, respectively denoted $V_I(x_I, x_E)$

[10] We follow here Reinganum (1983), which is itself an extension of Loury (1979) and Lee and Wilde (1980).

[e] A *Poisson process* is a stochastic process that is defined in terms of the occurrence of events (e.g., the arrival of customers in a simple queuing system or the number of web page requests at a server). Its main property is 'memorylessness': the number of arrivals occurring in any bounded interval of time after some date t is independent of the number of arrivals occurring before date t.

and $V_E(x_I, x_E)$.

$$V_I = \int_0^\infty e^{-rt} e^{-(h(x_I)+h(x_E))t} \cdot \left(\pi_0^m - x_I + h(x_I)\frac{\pi_1^m}{r} + h(x_E)\frac{\pi_{01}^d}{r}\right) dt$$
$$= \frac{\pi_0^m - x_I + (h(x_I)/r)\pi_1^m + (h(x_E)/r)\pi_{01}^d}{r + h(x_I) + h(x_E)},$$
$$V_E = \int_0^\infty e^{-rt} e^{-(h(x_I)+h(x_E))t} \cdot \left(h(x_E)\frac{\pi_{10}^d}{r} - x_E\right) dt = \frac{(h(x_E)/r)\pi_{10}^d - x_E}{r + h(x_I) + h(x_E)}.$$

Let us decompose the firms' expected profits. In the absence of an innovation, the incumbent realizes a profit of $(\pi_0^m - x_I)dt$. With 'probability' $h(x_I)dt$, the incumbent is first to innovate and derives a discounted profit (discounted to date t) of π_1^m/r. In these first two instances, the entrant does not make any profit. Finally, with 'probability' $h(x_E)dt$, the potential entrant is first to innovate and obtains π_{10}^d/r, leaving the incumbent with π_{01}^d/r.

A Nash equilibrium is a pair (x_I^*, x_E^*) such that x_j^* maximizes V_j^* given x_k^* $(j \neq k \in \{I, E\})$. Using a standard fixed-point argument, we can prove that a Nash equilibrium exists. Moreover, because of the strict concavity of h, we can assert that the Nash equilibrium is continuous in c and in π_0^m.

The question of interest is, of course, to compare x_I^* and x_E^*. The persistence of monopoly would obtain if $x_I^* > x_E^*$, meaning that the incumbent invests more in R&D than the entrant and is therefore more likely to be the first to innovate (and thus to remain a monopolist). Actually, the answer to this question depends on the combination of the two effects we mentioned above. On the one hand, the *efficiency effect* suggests that the incumbent has more incentives to innovate, as the net flow profit it receives by pre-empting the entrant (i.e., $\pi_1^m - \pi_{01}^d$) is larger than what the entrant gains by being first (i.e., π_{10}^d). On the other hand, the *replacement effect* is also present here because the marginal productivity of R&D expenditure for the incumbent decreases with its initial profits:

$$\frac{\partial}{\partial \pi_0^m}\left(\frac{\partial V_I}{\partial x_I}\right) < 0.$$

The intuition behind this finding is that, by increasing x_I, the incumbent moves the discovery date forward and hastens its own replacement. In contrast, the entrant does not forego a flow profit when innovating. It is not clear which effect will dominate.

Lesson 18.8 **In a patent race, it is in general ambiguous whether the incumbent or the entrant has a stronger incentive to invest.**

Yet, considering extreme situations, we can be sure that it is not always the same effect that dominates. If we consider a *non-drastic innovation* (as we have mainly done so far), making h almost linear would lead to a situation where the efficiency effect dominates: at equilibrium, the incumbent engages in more R&D than the entrant ($x_I^* > x_E^*$). In such a case, R&D is high, discovery is made early and the incumbent is concerned with the possibility of innovation by the entrant, whereas the replacement effect is not important. On the contrary, if we consider instead a *drastic innovation*, then we can prove that the replacement effect dominates: at equilibrium, the incumbent does *less* R&D than the entrant ($x_I^* < x_E^*$). Indeed, if the entrant discovers a drastic innovation, it becomes a de facto monopolist and hence, the efficiency effect disappears. This result lends credence to Schumpeter's conjecture about *creative destruction* (i.e., the idea that innovation leads one monopoly to replace another).

Case 18.2 illustrates the ambiguity mentioned in Lesson 18.8. It is impossible to say nowadays who will win the race to fuel the car of the future: one among the incumbent automobile manufacturers or some newcomer?

Case 18.2 The race for cleaner cars

The threats of global warming caused by the rising emissions of greenhouse gases exert an increasing pressure on automobile manufacturers to make cleaner cars. It is indeed estimated that surface transportation generates about a quarter of the man-made carbon emissions (and in this category, cars, buses and trucks are much larger pollution sources than ships and trains). In the traditional car industry, Toyota took the lead with its gasoline–electric hybrid, the Prius. Other carmakers have developed cars powered by hydrogen fuel cells (which make electricity by mixing the gas with oxygen), but mass production of such cars is not yet on the agenda. Despite its very large R&D spending, the car industry could well be unseated by disruptive technologies proposed by firms outside the industry. For instance, Tesla Motors, a Silicon Valley upstart (funded by prominent entrepreneurs such as the founders of PayPal and Google), has come up with a range of all-electric cars, powered by lithium-ion batteries and largely made of lightweight composites.

18.2.3 Socially excessive R&D in a patent race

As we will argue in the next chapter, there is a general presumption that markets provide too little incentive to introduce new innovations, which justifies legal protection of IP. It must be noted, however, that IP protection might sometimes go some steps too far by providing *too much* incentive to undertake R&D. Because of their 'winner-takes-all' nature, patent races might well lead to socially wasteful duplication of efforts.

To show this, we use a different (and simpler) model of a patent race. In the model of the previous subsection, uncertainty was about the time of invention and firms were choosing the intensity of R&D. We assume now instead that uncertainty is about the *success* of an invention and that firms choose the *scale* of R&D. In particular, suppose that two firms consider incurring a fixed cost of f to establish a research division, in the hope of finding a new product. This hope is translated by a probability of success equal to ρ. If only one firm finds the new product, it will obtain the monopoly profit π^m on the product market; this monopoly is guaranteed by a patent sufficiently broad to prevent imitation. If both firms find the new product, they will both offer it and each firm will obtain the duopoly profit π^d.

We want to determine the Nash equilibrium of the game in which the two firms simultaneously decide whether or not to invest in R&D. With the previous information, we can compute the expected profits depending on the firms' investment decisions. If a firm does not invest in

R&D, it does not enter the new market and its profit is therefore set to zero. If a firm does invest in R&D, its expected profit depends on whether the rival firm also invests or not. If the rival firm does not invest, then the firm active in R&D will incur the cost f and get the monopoly profit with probability ρ: its expected profit is thus given by $\rho\pi^m - f$. On the contrary, if the rival firm also invests in R&D, the firm will get the monopoly profit if it is the only one to be successful, or the duopoly profit in case both firms are successful; as the respective probabilities of these two events are $\rho(1-\rho)$ and ρ^2, each firm's expected profit is equal to $\rho(1-\rho)\pi^m + \rho^2\pi^d - f$. We easily see that the Nash equilibrium involves investment by both firms provided that

$$\rho(1-\rho)\pi^m + \rho^2\pi^d - f \geq 0 \Leftrightarrow f \leq \rho(1-\rho)\pi^m + \rho^2\pi^d \equiv f_2^{priv}.$$

Let us now adopt a public policy perspective and ask when it is socially optimal to have both firms investing in R&D. Our welfare measure is the sum of firms' profits and of consumer surplus. We define $W^m = \pi^m + CS^m$ as welfare in the monopoly case and $W^d = 2\pi^d + CS^d$ as welfare in the duopoly case. In general, consumers are better off if the marketplace is more competitive: $CS^d > CS^m$. From society's point of view, it is optimal to have one research division rather than two if

$$\rho W^m - f \geq \rho^2 W^d + 2\rho(1-\rho)W^m - 2f.$$

The right-hand side is the expected welfare if two divisions are active: with probability ρ^2, both divisions are successful and a duopoly situation ensues; with total probability $2\rho(1-\rho)$, only one division is successful and a monopoly situation ensues; in any case, the fixed cost f is paid twice. Rewriting the condition, we find that having just one research division is socially optimal if

$$f \geq \rho^2 W^d + \rho(1-2\rho)W^m \equiv f_2^{publ}.$$

Excessive R&D would occur if the Nash equilibrium of the investment game involved both firms investing (i.e., if $f \leq f_2^{priv}$) while society would find it optimal to have only one firm doing so (i.e., if $f \geq f_2^{publ}$). Excessive R&D therefore requires $f_2^{priv} > f_2^{publ}$, which can be rewritten as

$$\rho^2(\pi^m - \pi^d) > \rho^2(CS^d - CS^m) + \rho(1-\rho)CS^m. \tag{18.4}$$

On the left-hand side, we have the negative externality that a firm exerts on its rival when their R&D investments are successful: with probability ρ^2, profit is reduced from π^m to π^d. As this negative effect is ignored by firms but matters for society, it can lead firms to overinvest. Yet, as shown on the right-hand side of the inequality, an opposite force may lead firms to underinvest: in their private decision, firms ignore the positive effect that their successful investment has on consumer surplus when the other firm also invests; that is, if the other firm is successful (i.e., with probability ρ^2), welfare increases from CS^m to CS^d, whereas if the other firm is not successful (i.e., with probability $\rho(1-\rho)$), welfare increases from 0 to CS^m. In sum, we can draw the following lesson.

Lesson 18.9 **Imperfectly competitive firms tend to overinvest in R&D when their investment decreases the other firms' profit more than it increases consumer surplus.**

As an illustration, consider the following specification of a homogeneous-product market: the marginal cost of production is equal to zero and the inverse demand is given by $p = 1 - q$. If only one firm discovers the product, the profit-maximizing quantity is 1/2, yielding a profit of $\pi^m = 1/4$ and a consumer surplus of $CS^m = 1/8$. When both firms discover the product, suppose first that they compete in prices. Then, $\pi^d = 0$ and $CS^d = 1/2$. Plugging these values into condition (18.4), we find that the condition is never satisfied: in this model, firms never overinvest in R&D under Bertrand competition. In contrast, if firms compete à la Cournot, we have $\pi^d = 1/9$ and $CS^d = 2/9$. Then, condition (18.4) boils down to $\rho > 3/4$: in this model, firms overinvest in R&D under Cournot competition if the probability of success is large enough.

18.3 R&D cooperation and spillovers

In this section, we continue to consider innovative environments where ideas are common knowledge. That is, we have in mind industries such as automobiles in which new models are continually being developed, meaning that all firms in the industry have the simultaneous opportunity to achieve competing innovations. In terms of the modelization of the innovative process, we leave aside the tournament and uncertain aspects of innovation that we considered in Section 18.2 and come back instead to what we assumed in Section 18.1, that is, that R&D investments result immediately and for sure in an innovation. This simplification allows us to analyse more deeply a series of issues pertaining to the *strategic use of R&D* and to *R&D cooperation*. To this end, we depart from the setting used in Section 18.1 in several ways. First, we recognize that R&D is like any form of investment in that it precedes the production stage; as a result, we take into account the issues of strategic commitment that inevitably arise in considering R&D decisions. Therefore, we place firms in a symmetric position by allowing them all to invest in R&D (as is realistic when ideas are common knowledge rather than scarce). Second, we note that R&D exhibits many of the attributes of a public good (see Chapter 19) and, hence, we model the fact that R&D by one firm typically leads to *spillovers* which benefit other firms.[f] Finally, we incorporate in our setting the possibility for firms to cooperate on R&D decisions and, thereby, to internalize spillovers. Such a form of cooperation is indeed widespread and is widely allowed, if not encouraged, by public authorities (see Case 18.3 later). These different aspects of R&D (strategic behaviour, spillovers, cooperation) have been widely studied.[11]

We will base our discussion on a model that has the merits of allowing us to disentangle the separate influence of strategic behaviour and R&D cooperation, and to treat in a unified way

[f] Spillover levels vary drastically across industries. They are often inversely related to the level of patent protection. For instance, low-tech mature industries (e.g., paper) exhibit low effective patent protection and, hence, high spillovers; conversely, R&D-intensive industries (e.g., pharmaceutical drugs and software) exhibit high effective patent protection and low spillovers. See Griliches (1990).

[11] We follow here Leahy and Neary (1997) who provide a useful synthesis and extension of this literature, with the aim of disentangling the separate influence of strategic behaviour on the one hand and R&D cooperation on the other. The idea that R&D generates incentives for firms to behave strategically was first examined by Brander and Spencer (1983), assuming no R&D spillovers between firms and Cournot competition on the product market. Spence (1984) and Okuno-Fujiwara and Suzumura (1990) extended this analysis by considering, respectively, positive spillovers and Bertrand competition. Regarding R&D cooperation, d'Aspremont and Jacquemin (1988) presented a seminal analysis, which was then extended by, for example, Kamien, Muller and Zang (1992) and generalized by Amir, Evstigneev and Wooders (2003).

the cases where the firms' decisions on the product market are strategic substitutes or strategic complements.

We consider an industry of two symmetric firms ($i = 1, 2$), which compete in a two-stage game. At the first stage, firms simultaneously conduct process R&D and choose R&D expenditures $r(x_i)$ to reduce their marginal costs by x_1 and x_2, respectively. We assume that R&D activities exhibit decreasing returns to scale, that is, $r' > 0$ and $r'' > 0$ for x_i strictly less than $c/2$.[12] At the second stage, upon observing x_1 and x_2, firms compete in the product market with substitutable products. Denoting by σ_i firm i's strategic choice in the second stage, the analysis encompasses both quantity competition ($\sigma_i = q_i$) and price competition ($\sigma_i = p_i$).

We assume that marginal production costs are independent of output but decreasing in R&D, both that of the firm itself and, through spillover effects, of its rivals:

$$c_i(x_i, x_j) = c - x_i - \beta x_j, \tag{18.5}$$

where $\beta \in [0, 1]$ is the spillover coefficient. The parameter β measures the extent to which firm i benefits from R&D undertaken by firm j: when β is equal to zero, R&D is a private good as R&D expenditures benefit only the firm undertaking them; at the other extreme ($\beta = 1$), R&D is a pure public good as a firm fully benefits from the other firm's R&D.

We can write each firm's profits as follows:

$$\widetilde{\pi}_i = \pi_i[c_i(x_i, x_j), \sigma_i, \sigma_j] - r(x_i).$$

In the latter expression, π_i denotes the firm's net revenue from production and sales; it depends on the firm's unit production costs and on its own and rival firm's second-stage strategic choices.

Regarding competition on the product market, each firm's first-order condition sets equal to zero the partial derivative of its profit function with respect to its own action (i.e., either quantity, $\sigma_i = q_i$, or price, $\sigma_i = p_i$):

$$\frac{\partial \pi_1}{\partial \sigma_1} = 0 \quad \text{and} \quad \frac{\partial \pi_2}{\partial \sigma_2} = 0.$$

Assuming that the second-order condition is satisfied, that is,

$$\pi^{ii} \equiv \frac{\partial^2 \pi_i}{\partial \sigma_i^2} < 0,$$

we obtain a unique Nash equilibrium, which we denote $\{\sigma_1^*(x_1, x_2), \sigma_2^*(x_1, x_2)\}$, by solving the system of two first-order conditions.

As we know from Chapter 3, the sign of the cross-partial derivative of profit, denoted π^{ij}, depends on the nature of competition. We postulate that, under quantity competition, quantities are strategic substitutes, $\pi^{ij} < 0$, whereas, under price competition, prices are strategic complements, $\pi^{ij} > 0$. In sum,

$$\pi^{ij} \equiv \frac{\partial^2 \pi_i}{\partial \sigma_i \partial \sigma_j} = \frac{\partial^2 \pi_j}{\partial \sigma_i \partial \sigma_j} \begin{cases} < 0 & \text{under quantity competition,} \\ > 0 & \text{under price competition.} \end{cases}$$

This difference will prove crucial for assessing the effect of strategic behaviour and R&D cooperation.

[12] We also assume that $r(x)$ turns to infinity as x goes to $c/2$.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

18.3.1 Effects of strategic behaviour

By considering a two-stage game, we implicitly assume that firms are not able to commit to their second-stage choices (i.e., quantity or price) at the same time as they choose their R&D. As a consequence, investments in R&D are strategic in the sense that they are carried out with a view to affecting the environment in which the second-stage game is played.[g] To isolate this strategic incentive to invest in R&D, we contrast the results of the two-stage game with those of a hypothetic one-stage game in which firms choose simultaneously their R&D intensity and their product market actions.

We focus here on the case where firms do not cooperate at the R&D stage (i.e., firms choose their R&D expenditure independently). Hence, firm i chooses x_i to maximize its first-stage profit (which incorporates the second-stage equilibrium):

$$\widetilde{\pi}_i(x_i, x_j) = \pi_i[c_i(x_i, x_j), \sigma_i^*(x_i, x_j), \sigma_j^*(x_i, x_j)] - r(x_i).$$

The first-order condition for profit maximization is given by

$$\frac{d\widetilde{\pi}_i}{dx_i} = 0 \Leftrightarrow \underbrace{\frac{\partial \pi_i}{\partial c_i}\frac{\partial c_i}{\partial x_i}}_{\text{direct effect}} + \underbrace{\frac{\partial \pi_i}{\partial \sigma_i}\frac{\partial \sigma_i^*}{\partial x_i}}_{=0} + \underbrace{\frac{\partial \pi_i}{\partial \sigma_j}\frac{d\sigma_j^*}{dx_i}}_{\text{strategic effect}} = r'(x_i).$$

Let us decompose the latter condition. When firm i assesses the effect of increasing its R&D intensity, it first considers the *direct* or *'cost-minimizing' effect* a further cost reduction will have on its profit ($\partial \pi_i / \partial x_i$). From the production costs (18.5), we have that the direct effect is simply equal to the equilibrium second-stage quantity q^*. If firm i were non-strategic, this is the only effect that would matter for its choice. We can thus say that *in the absence of strategic behaviour and of R&D cooperation, the marginal private return to R&D per unit of output is simply the reduction in the firm's own unit costs* (i.e., 1 under our assumptions).

However, when firm i is strategic, it also anticipates the effect of its R&D choice on the subsequent product market equilibrium. There are two potential effects. A first effect should be ignored; indeed, using the envelope theorem, we know that the effect on π_i from the change in firm i's own second-stage action is nil (since σ_i is chosen so that $\partial \pi_i / \partial \sigma_i = 0$). Alternatively, the second effect is crucial. This is the *strategic effect* resulting from the combined influence of firm i's investment on firm j's second-stage action ($d\sigma_j^* / dx_i$) and of firm j's action on firm i's profit ($\partial \pi_i / \partial \sigma_j$).

As for the latter derivative, it is easy to see that firm i is hurt when firm j increases its quantity, but benefits when firm j increases its price. That is, $\partial \pi_i / \partial \sigma_j < 0$ if $\sigma_j = q_j$, while $\partial \pi_i / \partial \sigma_j > 0$ if $\sigma_j = p_j$. Assessing the sign of the former derivative is trickier as an increase in x_i reduces not only firm i's marginal cost but also firm j's (unless spillovers are nil). As both firms become tougher competitors, we understand that the net effect on firm j's second-stage decision (σ_j^*) will depend on the nature of the strategic variables and on the degree of spillovers. We thus treat quantity and price competition separately.

Let us start with quantity competition. We know from Chapter 3 that when firms produce substitutable products, quantity competition typically refers to a situation with strategic substitutability, which implies downward-sloping reaction functions. As depicted in the left panel of

[g] In this respect, the present model is reminiscent of the one we used in Chapter 16 to elaborate a taxonomy of entry-related strategies.

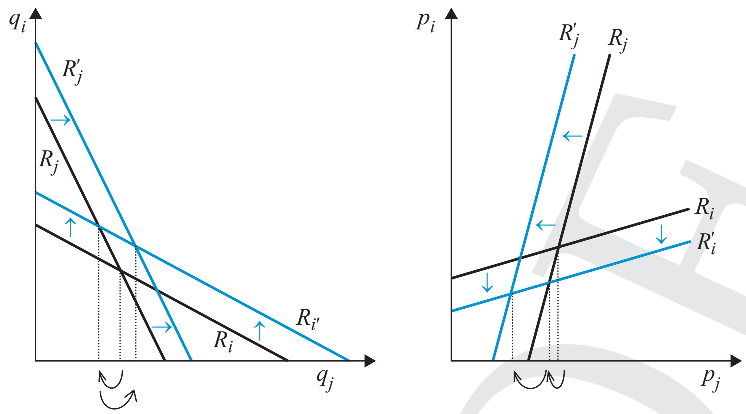

Figure 18.3 Strategic effect of R&D investments (left: quantity competition; right: price competition)

Figure 18.3, an increase in x_i allows firm i to move its reaction function to the right (from R_i to R'_i): because firm i has a lower marginal cost, it reacts to any firm j's quantity by producing a larger quantity than before. In the absence of spillovers ($\beta = 0$), the analysis stops here: firm j's reaction function does not move and the new equilibrium is such that firm j produces a lower quantity as a result of the increase in x_i. However, for $\beta > 0$, firm i's R&D investment also reduces firm j's marginal cost, which shifts firm j's reaction function to the right (from R_j to R'_j). As shown in the figure, if firm j's reaction function moves sufficiently outwards (i.e., if spillovers are large enough), the new equilibrium is such that firm j produces a *larger* quantity than before. There exists thus a threshold value of the spillover parameter around which the sign of the strategic effect changes. We denote this threshold by $\bar{\beta}$. We show below how to derive the value of $\bar{\beta}$ in our general model. For now, we observe:

$$\text{If } \sigma_j = q_j, \text{ then } d\sigma_j^*/dx_i < 0 \text{ for } \beta < \bar{\beta} \quad \text{and} \quad d\sigma_j^*/dx_i > 0 \text{ for } \beta > \bar{\beta}.$$

From this, we conclude that the strategic effect of an increase in the R&D of one firm on its own profit is positive for small spillovers ($\beta < \bar{\beta}$) and negative for large spillovers ($\beta > \bar{\beta}$).

We repeat now the analysis for price competition. Because of strategic complementarity, reaction functions slope upwards. A decrease in a firm's marginal cost allows this firm to set a lower price for any price of the rival firm, that is, to shift its reaction functions inwards (from R_i to R'_i). As illustrated in the right panel of Figure 18.3, we now have two reinforcing effects: (i) by reducing c_i, the increase in x_i shifts firm i's reaction function down and brings the equilibrium towards a lower value of p_j; (ii) by also reducing c_j (when $\beta > 0$), the increase in x_i shifts firm j's reaction function to the left (from R_j to R'_j), which decreases further firm j's equilibrium price. Hence, we conclude:

$$\text{If } \sigma_j = p_j, \text{ then } d\sigma_j^*/dx_i < 0 \text{ for all values of } \beta.$$

Recalling that $\partial\pi_i/\partial\sigma_j < 0$ if $\sigma_j = q_j$ and $\partial\pi_i/\partial\sigma_j > 0$ if $\sigma_j = p_j$, we are now in a position to sign the strategic effect.[h]

[h] It is possible to show that under fairly general conditions, $\bar{\beta}$ lies between 0 and 1 under quantity competition, and $\bar{\beta} < 0$ under price competition.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

Lesson 18.10 **The strategic effect of an increase in the R&D of one firm on its own profit is (i) positive for small spillovers ($\beta < \bar{\beta}$) and negative for large spillovers ($\beta > \bar{\beta}$) under quantity competition; (ii) always negative under price competition.**

Using the terminology introduced in Chapter 16, we can summarize the intuition behind this result as follows. Because an increase in its R&D expenditure makes the firm a tougher competitor, it is worth investing more from a strategic point of view only if tough behaviour is met by a soft response of the rival firm. This is the case under quantity competition, provided that spillovers are small enough (because otherwise, the other firm also becomes a tougher competitor), and is never the case under price competition. As a corollary, if the rival reacts toughly (i.e., under price competition or under quantity competition with strong spillovers), strategic firms choose optimally to invest less in R&D than they would do were they only motivated by cost minimization. In that case, strategic behaviour leads to larger marginal costs and thus, to lower output.

18.3.2 Effects of R&D cooperation

Suppose now that firms cooperate in their choice of R&D levels (i.e., they choose them to maximize joint profits), though they continue to compete at the second stage. For the moment, we assume that cooperation does not affect the value of the spillover parameter (we will relax this assumption below).[13] In this case, the first-order condition for joint profit maximization in the first stage is given by

$$\frac{d(\widetilde{\pi}_i + \widetilde{\pi}_j)}{dx_i} = 0 \Leftrightarrow \left| \begin{array}{l} \underbrace{\frac{\partial \pi_i}{\partial c_i}\frac{\partial c_i}{\partial x_i}}_{\text{direct effect}} + \underbrace{\frac{\partial \pi_i}{\partial \sigma_i}\frac{\partial \sigma_i^*}{\partial x_i}}_{=0} + \underbrace{\frac{\partial \pi_i}{\partial \sigma_j}\frac{d\sigma_j^*}{dx_i}}_{\text{strategic effect 1}} \\ + \underbrace{\frac{\partial \pi_j}{\partial c_j}\frac{\partial c_j}{\partial x_i}}_{\text{spillover effect}} + \underbrace{\frac{\partial \pi_j}{\partial \sigma_i}\frac{d\sigma_i^*}{dx_i}}_{\text{strategic effect 2}} + \underbrace{\frac{\partial \pi_j}{\partial \sigma_j}\frac{\partial \sigma_j^*}{\partial x_i}}_{=0} = r'(x_i)\,. \end{array} \right. \tag{18.6}$$

Because x_i is now chosen to maximize total profits, its effect on firm j's profit has also to be taken into account. There are three effects. First, because of spillovers, an increase in x_i affects directly firm j's profit by decreasing firm j's marginal cost. Obviously, this positive 'spillover effect' increases with the spillover parameter β. Second, a change in x_i modifies firm i's second-stage action, which in turn affects firm j's profits. One understands intuitively that this strategic effect is negative whatever the nature of competition: by investing more in R&D, firm i gains a competitive advantage over its rival; that is, firm i is able to produce more or to set a lower price in the second stage, which hurts firm j. One also understands that this negative strategic effect weakens when spillovers get stronger (since the competitors' efficiency is also enhanced). Finally, a change in x_i also affects firm j's equilibrium second-stage decision and, thereby, its profit. Yet, as firm j makes an optimal decision in the second stage, this effect can be ignored.

In sum, R&D activities in the presence of spillovers create two types of externality. The first externality affects overall industry profits and increases with the level of spillovers; it is

[13] See the seminal analysis of d'Aspremont and Jacquemin (1988).

ignored when firms choose their R&D levels separately but is internalized when firms choose their R&D levels so as to maximize their joint profits. The second externality affects a firm's competitive advantage with respect to its rival: firms invest in R&D to become relatively more efficient than their competitors. This externality, which weakens as spillovers increase, is present when firms choose their R&D investments separately but is fully internalized when they act cooperatively. We can therefore conclude that there exists a pivotal spillover rate above which the total effect of the two externalities is positive. Indeed, if spillovers are large enough, the competitive advantage motivation for investing in R&D is weak, whereas the temptation to free-ride on the other firm's effort is high; as a result, cooperation leads to larger investments in R&D, implying further reductions in unit costs and a larger output.

Lesson 18.11 **When firms behave strategically, R&D cooperation leads to more R&D when spillovers are large but to less R&D when spillovers are small.**

Note that when firms do not behave strategically (i.e., when they commit to both R&D levels and second-stage actions), only the direct and the spillover effect remain in expression (18.6). It follows that the effect of x_i on total profits is equal to $(1+\beta)\,q^*$. Hence, the private and social returns to R&D coincide: cooperation fully internalizes the externality arising from R&D spillovers and is thus desirable.

As an illustration, suppose that the second stage is a Cournot duopoly and that the inverse demand is given by $p = a - q_i - q_j$. Firm i chooses q_i to maximize $\pi_i = (a - c_i - q_i - q_j)q_i$. Solving for the first-order condition, we find firm i's reaction function: $q_i(q_j) = (1/2)(a - c_i - q_j)$. By analogy, firm j's reaction function is $q_j(q_i) = (1/2)(a - c_j - q_i)$. The equilibrium quantity of firm j is then easily found as $q_j^* = (1/3)(a - 2c_j + c_i)$, and its profits are simply equal to the square of the equilibrium quantity. As $c_i = c - x_i - \beta x_j$ and $c_j = c - x_j - \beta x_i$, we can rewrite firm j's equilibrium profit as

$$\pi_j^* = \tfrac{1}{9}(a - c + (2-\beta)x_j + (2\beta - 1)x_i)^2.$$

When it comes to choose x_i in the first stage, R&D cooperation leads to a larger value provided that an increase in x_i implies an increase in π_j^*. We see from the above expression that the condition for this is $2\beta - 1 > 0$. Hence, in this particular case, $\bar{\beta}' = 1/2$.

R&D cooperation and information sharing

We refer to the previous mode of R&D cooperation as the formation of an *R&D cartel*. One can distinguish it from the case of a *cartelized research joint venture (RJV),* in which firms not only coordinate their R&D decisions but also share their information completely so as to eliminate duplication of effort. As a result, in a cartelized RJV, the final spillover parameter β is internally set to unity. This naturally tends to make cooperation more attractive from a welfare point of view. Indeed, it turns out that *a cartelized RJV yields a superior performance compared with*

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

a non-cooperative R&D in all criteria of interest: propensity for R&D, firms' profits, consumer surplus and thus social welfare.[14]

The previous theoretical analysis leads thus to unambiguous antitrust implications: *public authorities should permit simultaneous R&D sharing and coordination of R&D decisions among firms that compete in a product market*. No direct action seems to be needed to encourage such cooperation as the firms' incentives for cooperation in R&D are clear (information sharing and coordination of R&D decisions yield higher profits); public authorities just need to provide the attending legal framework for such cooperative arrangements, which corresponds to what is currently done in the USA, the EU and Japan, as described in Case 18.3.

Case 18.3 Antitrust provisions related to R&D cooperation

As we will make clear in the next chapter, IP-generating activities suffer from three sources of market failure: externalities, indivisibilities and uncertainty. When it comes to R&D, one way to alleviate these three problems is to allow firms to form a cooperative R&D venture. As we have emphasized in this section, rival firms often exert positive externalities on one another through their research activities. Indeed, new knowledge easily spreads across firms, meaning that firms freely benefit (at least partly) from the R&D efforts of their rivals. As we have seen, these knowledge spillovers raise the prospect of free-riding and firms might be inclined to cut back their R&D spending. In response to this problem, the formation of a cooperative R&D venture allows firms to internalize the externalities and, thereby, preserve their incentives to do R&D.

Cooperative R&D ventures are also a way to pool risk and, hence, to better manage technological and market uncertainty. Furthermore, cooperative R&D ventures might also reduce the problems stemming from indivisibilities by allowing firms to share costs, eliminate useless duplication of R&D projects, pool complementary skills and exploit economies of scale.

Although cooperative R&D may be subject to contractual hazards (opportunistic behaviour, free-riding, difficulties about sharing the results, etc.), it is generally accepted that their formation is welfare-enhancing. This is why antitrust authorities tolerate this type of collaborative arrangement between firms.

In the USA, the National Cooperation Act passed in 1984 allows firms to cooperate in R&D provided they remain competitors on product markets. A similar permissive antitrust attitude towards R&D cooperation is the norm in Japan and in Europe. In 2010, the European Commission considerably extended the scope of the R&D *Block Exemption Regulation*, so as to exempt from the application of Article 101 of the Treaty on the Functioning of the European Union (which prohibits horizontal agreements among firms) not only R&D activities carried out jointly but also agreements where one party finances the R&D activities carried out by the other party. Furthermore, public policies, such as the European Framework Programmes, explicitly encourage firms to pool their R&D activities.

[14] See Amir, Evstigneev and Wooders (2003), who generalize the results of Kamien, Muller and Zang (1992).

18.3.3 Further analysis of R&D cooperation

The previous analysis can be extended in several directions, essentially with respect to (i) the nature of R&D spillovers, (ii) the design of R&D cooperation and (iii) the potential effect of R&D cooperation on product market collusion. We briefly review here a number of those extensions.

The nature of R&D spillovers

In the setting we used, knowledge spillovers were modelled as a 'manna from heaven': firms were automatically benefiting from the other firms' R&D effort (according to the factor β). The realism of this assumption can be questioned. R&D may not only generate new information, but also enhance the firm's ability to identify, assimilate and exploit existing information from the environment. That is, R&D also contributes to develop a stock of prior knowledge, which is called *absorptive capacity*,[15] and thereby increases a firm's ability to learn from others. In the strategic games we have considered above, introducing this additional property of R&D has two opposite effects on R&D investments: on the one hand, the firm wants to learn more from the rival firms and therefore has an incentive to raise its own R&D expenditure; yet, on the other hand, this will increase the other firms' incentive to free-ride, meaning that there will be less to learn from.

To formalize this intuition we rewrite the cost function (18.5) as follows:[16]

$$c_i(x_i, x_j) = c - x_i - B(x_i)x_j,$$

where $0 \leq B(x_i) \leq 1$ describes the proportion of R&D that spills over from firm j to firm i. In contrast with our previous formulation, the spillover rate is no longer exogenously fixed ($B(x_i) = \beta$) but is an increasing function of the firm's own investment: we assume that $B'(x_i) > 0$ to capture the idea that by investing more, firm i increases its absorptive capacity and, thereby, its ability to learn from firm j.

To evaluate the impact of this modification, let us consider a Cournot duopoly for a homogeneous product, whose inverse demand is given by $p = a - q_i - q_j$ (with $a > c$).[17] Assume also that R&D costs are given by $r(x_i) = (1/2)x_i^2$. In the second stage of the game (production stage), firm i chooses q_i to maximize

$$\pi_i = (a - q_i - q_j)q_i - (c - x_i - B(x_i)x_j)q_i.$$

From the first-order condition, we derive firm i's reaction function (where b stands for $a - c$): $q_i(q_j) = \frac{1}{2}(b - q_j + x_i + B(x_i)x_j)$. Proceeding in a similar way for firm j and solving for the system of the two reaction functions, we find the Nash equilibrium of the production stage as

$$q_i(x_i, x_j) = \tfrac{1}{3}[b + (2 - B(x_j))x_i + (2B(x_i) - 1)x_j].$$

Equilibrium profits are

$$\widetilde{\pi}_i(x_i, x_j) = [q_i(x_i, x_j)]^2.$$

Consider now the first stage of the game (R&D stage) in the absence of cooperation. Firm i's problem is to choose the R&D level x_i that maximizes $\widehat{\pi}_i(x_i, x_j) = \widetilde{\pi}_i(x_i, x_j) - (1/2)x_i^2$.

[15] See Cohen and Levinthal (1989).

[16] We follow the analysis of Grünfeld (2003). For an alternative approach, see Kamien and Zang (2000).

[17] This is the basic setup of d'Aspremont and Jacquemin (1988).

Using the above expressions, we can write the first-order condition as: $\partial\widehat{\pi}_i/\partial x_i = 0$ if and only if

$$\tfrac{2}{9}(b + (2 - B(x_j))x_i + (2B(x_i) - 1)x_j)(2 - B(x_j) + 2B'(x_i)x_j) = x_i.$$

Assuming that firms are symmetric, we have $x_i = x_j = x$ at the symmetric equilibrium and we can rewrite the previous expression as

$$\begin{aligned} \frac{\partial\widehat{\pi}}{\partial x} &= \tfrac{2}{9}\left(b + (1 + B(x))\,x\right)\left(2 - B(x) + 2B'(x)\,x\right) - x = 0 \\ &\iff Vx^2 + (W - Y)\,x + Z = 0, \end{aligned} \tag{18.7}$$

where $V \equiv 4\,(1 + B(x))\,B'(x)$, $W \equiv 4bB'(x)$, $Y \equiv 9 - 2\,(1 + B(x))\,(2 - B(x))$ and $Z \equiv 2b\,(2 - B(x))$. Because $0 \leq B(x) \leq 1$ and $B'(x) > 0$, we have $V, W, Y, Z > 0$.

We want to compare the equilibrium R&D investment level in the game where the spillover rate is affected by absorptive capacity effects ($B(x)$) to the game of the previous section where the spillover rate is exogenous (β). Let us denote the former equilibrium R&D level by x_a (with a for absorptive) and the latter by x_e (with e for exogneous). Using the first-order condition (18.7), we have that x_a and x_e are respectively defined by

$$V_a x_a^2 + (W_a - Y_a)\,x_a + Z_a = 0 \quad \text{and} \tag{18.8}$$

$$-Y_e x_e + Z_e = 0, \tag{18.9}$$

where the second line follows from the fact that $B'(x) = 0$ if the spillover rate is exogenous (implying that $V_e = W_e = 0$). We can now show that if the spillover rate ($B\,(x_a)$) generated by the game with absorptive capacity effects is the same as the exogenous spillover rate (β) in the game of the previous section, then we necessarily have that $x_a > x_e$: *the presence of absorptive capacities induces larger R&D investments for a similar (ex post) spillover rate*. The proof goes as follows. If $B\,(x_a) = \beta$, then $Y_e = Y_a$ and $Z_e = Z_a$. Combining (18.8) and (18.9), we have $V_a x_a^2 + (W_a - Y_a)\,x_a + Y_a x_e = 0$, or $(V_a x_a + W_a)\,x_a = Y_a\,(x_a - x_e)$. As $(V_a x_a + W_a)\,x_a > 0$ and $Y_a > 0$, it necessarily follows that $x_a > x_e$. Intuitively, by imposing $B\,(x_a) = \beta$, we 'freeze' the traditional negative effect of spillovers on R&D investments (i.e., the free-riding effect) and we are left, in the model with absorptive capacity, with a pure *positive learning effect* of own R&D that drives up the incentive to invest.

Design of R&D cooperation

One shortcoming of the approach we followed above is that it assumed away the question of the endogenous formation of research joint ventures. Indeed, for simplicity, we have restricted attention to a duopoly. Results do not hinge critically on this assumption and one can consider an arbitrary number of firms. However, a limitation of any such analysis is that we have compared complete non-cooperation in R&D with industry-wide R&D cooperation.[18] In many industries, (i) there are typically several research joint ventures competing with each other, or (ii) firms' R&D cooperation is bilateral and non-exclusive in nature (translating into situations where firms i and j, and j and k collaborate with one another, respectively, while firms i and k do not collaborate). To account for the former possibility, we can model the formation of RJVs in terms of a *coalition structure*, which is a partition of the set of firms (each firm belonging to one RJV only). To account

[18] See Kamien, Muller and Zang (1992) and Leahy and Neary (1997).

for the latter possibility, we need to consider two-player relationships, which collectively generate a network structure.[19]

R&D cooperation and product market collusion

So far we have assumed implicitly that firms act non-cooperatively on the product market. However, a major antitrust concern is the possibility that cooperation in R&D could pave the way for, or at least increase the likelihood of, collusive behaviour in the product market. If this conjecture turns out to be true, we might have to reconsider the policy recommendation according to which R&D cooperation and information sharing should be permitted, if not encouraged. Indeed, if R&D cooperation appears to worsen product market performance, then public policy towards R&D cooperation involves a trade-off between market power and efficiency, and it is by no means clear that this trade-off will always balance out in favour of R&D cooperation. As indicated in Case 18.4, there are reasons to fear collusive behaviour on the part of firms forming R&D joint ventures.

Case 18.4 Research joint ventures and collusion[20]

May collusive behaviour be facilitated by RJVs? Answering this question empirically is extremely difficult as it requires one to isolate the impact of the returns to collusion on the decision to join an RJV from the other factors determining the decision to join the venture. Sovinsky and Helland (2012) have found a clever way around this difficulty. They exploit the variation in RJV formation generated by a quasi-experiment that affects the collusive benefits of an RJV while not directly affecting the research synergies associated with that venture. Their identification strategy is simple: if product market collusion is not a motivation to form an RJV then, after controlling for firm, RJV and industry characteristics, the propensity to enter into an RJV should not be impacted by changes in the antitrust policy aimed at deterring collusion in the final goods market.

The change in antitrust policy that the authors consider took place in 1993 in the USA: at that date, the revision of the so-called ‘leniency policy programme’ made it more attractive for cartel members to report illegal behaviour, thereby making collusion harder to sustain (see Chapter 14). The authors apply their analysis to three industries: petroleum manufacturing, computer and electronic product manufacturing and telecommunications. These industries share two characteristics: RJV participation is very high and there is a history of antitrust suits.

After controlling for industry characteristics, RJV characteristics, firm attributes and correcting for the endogeneity of R&D, the authors find that, among telecom firms, the decision to join an RJV is impacted by the policy change in a very significant way: the revised leniency policy reduces the probability that telecom firms join a given RJV by 34%. Similar reductions are observed for computer and semiconductor manufacturers (33%) and for firms in petroleum refining (27%).

[19] See Yi and Shin (2000) for the coalition approach, and Goyal and Moraga-González (2001) for the network approach.
[20] This case is based on Sovinsky and Helland (2012).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

Review questions

1. Why does a monopolist have fewer incentives to innovate than a perfectly competitive firm? Explain the meaning of the 'replacement effect'.
2. Why does a monopolist threatened by entry have more incentives to innovate than a potential entrant? Explain the meaning of the 'efficiency effect'.
3. Explain why firms might invest too much in R&D (from a social point of view) when they are racing to obtain a patent on an innovation.
4. Discuss the effects of strategic behaviour on firms' investments in R&D.
5. Is it a sensible policy to allow firms to coordinate their R&D decisions? Discuss.

Further reading

An important article to read is the seminal article of Arrow (1962), in which he explores the problems associated with activities generating new knowledge and the influence of market structure on the incentives to innovate. For the reverse influence of innovation on market structure, see Gilbert and Newbery (1982) and their explanation of the 'efficiency effect' leading to the persistence of monopoly. Our treatment of asymmetric patent races was based on Reinganum (1983). As for the effect of strategic behaviour and cooperation on R&D decisions, the seminal paper is d'Aspremont and Jacquemin (1988). For a unified treatment of these issues (incorporating strategic substitutes and complements on the product market), see Leahy and Neary (1997).

Exercises

18.1 Incentives to R&D and market structure

At a hotel in Munich in May 2007, Webasto, a German auto parts maker, decided to license the rights to one of its best-selling products – a rooftop solar panel for cars and trucks – to the highest bidder at a public auction (see *International Herald Tribune*, 13 May 2007). Firms coming from different industries attended this public auction. From which type of industry do you think the highest bidder for Webasto's products came? From a concentrated (monopoly-like) industry? From a much more competitive industry with numerous small players? Or from some intermediate (oligopoly-like) industry? The following exercise will help you answer these questions.

Assume that the demand for trucks is $p = 100 - q$ (where q is the quantity and p is the price), and that Webasto's rooftop solar panel allows truck manufacturers to reduce the constant marginal cost of production from 70 to 60.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

1. Confirm that this is a non-drastic (or minor) innovation and that marginal cost would have to be reduced to less than 40 for the innovation to be drastic (or major).
2. Suppose that the industry is a *monopoly* (not threatened by entry). How much is this firm willing to pay (per period of time) to acquire the innovation?
3. Suppose that the industry is a *Bertrand oligopoly*. That is, there are n firms (with $n \geq 2$) that compete in price. Before the innovation, all firms have the same marginal cost of 70. After the innovation, one of them has a lower cost of 60. Compute how much the latter firm is willing to pay for the innovation.
4. Now assume that the market is served by *Cournot duopolists* who have identical marginal costs of 70 before the innovation.
   (a) Confirm that the pre-innovation price is 80 and that at this price each firm has profits per period of 100.
   (b) Suppose that one of these firms is granted use of the innovation. Confirm that the price falls to 76.67, and compute the per-period profits of the two firms.
   (c) How much is any of these duopolists willing to pay to acquire the innovation?
5. Suppose that the industry is a *monopoly threatened by entry*. More precisely, with the existing technology, production at a marginal cost of 70 does not make entry profitable. However, by lowering the marginal cost to 60, the new technology makes entry profitable. So, by acquiring the innovation, the incumbent firm precludes entry: it remains a monopolist and produces now at a marginal cost of 60. On the contrary, if the monopolist does not acquire the innovation, another firm does, which allows it to enter the market. The market structure thus becomes an asymmetric Cournot duopoly in which the incumbent firm has a marginal cost of 70, while the entrant has a marginal cost of 60.
   (a) How much is the incumbent firm willing to pay for the innovation?
   (b) How much is the entrant willing to pay for the innovation?
   (c) If the innovation goes to the highest bidder, what is the influence of innovation on market structure? Discuss.
6. Finally, by collecting your answers to questions (1) to (5), rank the various market structures according to the incentives to innovate that they convey to firms. Comment on your ranking.

18.2 Incentives to invest in product and process innovations

Consider the following duopoly. Each firm i ($i = 1, 2$) incurs a constant marginal cost equal to c_i and produces a differentiated product, q_i, sold at price p_i. The demand system is obtained from the optimization problem of a representative consumer. We assume a quadratic utility function which generates the linear inverse demand schedule $p_i = a - q_i - \gamma q_j$ in the region of quantities where prices are positive. The parameter $\gamma \in [0, 1]$ is an inverse measure of the degree of product differentiation: the lower γ, the more products are differentiated (if $\gamma = 1$, products are perfect substitutes; if $\gamma = 0$, products are perfectly differentiated). Firms compete à la Cournot on the product market.

Initially, both firms produce at cost $c_i = c$. A new process innovation allows firms to reduce the constant marginal cost of production from c to $c' = c - x$ (with $0 < x < c$). We assume that the innovation is non-drastic. That is, the cost reduction does not allow the innovator to behave like a monopolist. A sufficient condition is that the monopoly price corresponding to c' is larger than the initial cost c; that is, $(a + c - x)/2 > c$. Equivalently, assuming without loss of generality that the difference $a - c$ is equal to unity, we assume: $x < a - c = 1$.

1. Compute how much a duopolist is willing to pay for acquiring the innovation and being its single user (i.e., compute the difference between the profit a firm makes when it is the sole user of the innovation and the profit it makes when no firm uses the innovation).

2. Suppose now that a product innovation allows firms to increase product differentiation (i.e., to reduce the parameter γ). Show that the adoption of this product innovation *raises* the incentives to adopt the process innovation (computed at the previous question) if and only if the initial degree of product substitution, γ, is lower than $2/3$.

18.3 Strategic patenting

Consider a market where demand is given by $P(q) = a - q$. An incumbent firm has a proprietary technology with a constant marginal cost of c_I (with $c_I < a$). One other firm could enter the market as a Cournot duopolist, but the technology available to this firm does not allow it to make any positive profit if it enters. Precisely, the marginal cost corresponding to the entrant's technology, c_E, is such that $c_E = (a + c_I)/2 = \tilde{c}$.

1. Check that this condition implies the non-positivity of the entrant's quantity at the Cournot–Nash equilibrium.

Suppose now that alternative technologies become available with a constant marginal cost c comprised between c_I and $\tilde{c}$.

2. Show that, although the incumbent has no incentive to switch to any of these technologies, it has a higher incentive to acquire a patent on them than the entrant has.

3. What does the previous result tell you about firms' motivations to file patents? Discuss.

18.4 Essay question: 'Pay-for-delay' and generic drugs

Under the 'pay-for-delay' deal, the patent holder of a drug pays a maker of generic drugs to delay its launch of a cheap copy.

1. Discuss the likely consequence for consumer welfare of such behaviour.

2. Model the market for drugs as a homogeneous Cournot market with linear demand (it would be straightforward to include asymmetries between different pharma companies but symmetry makes the analysis particularly easy). Are there any contracts between the branded drugmaker and the generic drugmaker that lead to higher profits of both firms?

3. Some branded drugmakers have decided to offer lower-priced 'authorized' generic versions shortly before the patent expires. Note that the first (non-authorized) generic producer entering the market is protected from further competition for a period of

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
18 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 15:11

6 months. In light of this regulation, discuss the conseqences of authorized generic versions on entry and competition.

4. Some generic drugmakers now agree to delay the launch of the generic in return for a promise by the branded drugmaker not to launch an 'authorized' generic version itself. Discuss the likely consequences of such agreements on consumer and total welfare.

19 Intellectual property

In the previous chapter, we focused on the positive aspects of R&D by examining the interplay between market structure and innovation. In consequence, we were not too specific about the exact regime of IP protection. In this chapter, we want to adopt a more normative point of view and study how IP protection should optimally be organized. This chapter also provides the reader with a broad description of the realm of IP.

In Section 19.1, we study the link between innovation and IP in an essentially non-technical way. We describe the appropriability problem of innovation and we consider several ways to close the wedge that this problem drives between social and private rates of return from innovation. We start with the main policy instrument that has been designed to promote innovation, namely the institution of intellectual property and its legal protection (through IP rights such as patents and copyrights). We explain that the main rationale of IP rights is to provide incentives to produce information and knowledge by conferring a monopoly right to the producer. We then compare this institution to other public and private responses (namely rewards and secrecy), and we examine which of these solutions innovators tend to choose in practice.

In Section 19.2, we address some of the previous issues in a more formal way. First, we study the optimal design of IP rights. We start by arguing that the negative impacts that monopolies have on welfare call for limitations on the legal protection conferred by IP rights. What types of limitation? We examine two adjustable dimensions: the length and the breadth. The *length* simply refers to the duration of the protection, while the *breadth* refers to the usage the innovator can make of his rights with respect to competitors. Second, we compare the relative merits of rewards, secrecy and patents as mechanisms to encourage innovation and diffusion.

In the first two sections, we treat innovations as if they were produced in complete isolation. This simplifying assumption may be a good approximation for some industries but not for others, especially those where the technology is complex and where innovations result more from a *cumulative process*. In Section 19.3, we discuss the problems that cumulative innovations face when it comes to encouraging and diffusing them. One such problem is that a patent on a first-generation innovation confers the patentee a holdup right over subsequent innovations, whose development is based on the first one. A second problem arises for innovations that have to be combined to provide a result; here, the allocation of IP rights to separate right-holders results in higher total prices, thereby reducing access and usage. We will see that IP rights cannot solve these two problems unless they are complemented by additional mechanisms.

While the first three sections are centred around theoretical arguments, the last section has a more applied nature. In Section 19.4, we apply the previous analyses to the specificities of the digital economy. Two topical issues are addressed: the piracy of digital products and the protection of software.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

19.1 Innovation and IP: basics

We start this section by describing the appropriability problem that mars the production of information and knowledge. We consider then public and private responses to this problem. The main public responses consist either of restricting the exploitation of knowledge (by instituting IP and protecting it), or raising the expected returns of new knowledge by lowering its cost of production (through subsidization of research and patronage of artists, or by allowing putative competitors to form cooperative R&D ventures, as already described in the previous chapter). We examine these various responses in turn and discuss their respective merits. We close this section by looking at how innovators protect their R&D investments in practice. It turns out that the legal protection of IP is often complemented by private measures.

19.1.1 Information and appropriability

What do inventions, business methods, industrial processes, novels, songs, paintings, and so on have in common? They all result from the production of *information* (or *knowledge*). The problem with activities generating information is that they suffer from the three generic sources of market failure: uncertainty, indivisibilities and externalities. As a result, these activities face a generic problem of *appropriability*, which sets them apart from other investments made by firms or individuals.

A first source of market failure in information markets is *uncertainty*. Investments in R&D involve two types of uncertainty: on top of technological uncertainty (how to make new things and how to make them work), there is also commercial uncertainty (how to make new things adopted by the consumers). For instance, it is estimated that in the USA, the odds of a potential medicine making it through the R&D pipeline and being approved for patient use are about one in 10 000.[21] The same uncertainties apply for the creation of artistic and literary work. The decisions to produce or invest in information-generating activities are therefore necessarily mixed with decisions to bear risk. Separating the two types of decision is often difficult because of moral hazard: it is very hard to attribute the failure of a project to bad luck or to a lack of effort (which is typically unobservable). Hence, a balance has to be found between incentives and the transfer of risk, which undermines the efficiency of the investment.

Indivisibilities are a second source of failure in information markets. The creation of new knowledge and new information involves large fixed setup costs. For instance, the innovation cost of a new molecule in the pharmaceutical industry is estimated to be 1.25 billion euros; the remake of *King Kong* by Peter Jackson cost 170 million euros. As such activities often require the division of highly specialized labour, they are also prone to economies of scale. Finally, both knowledge and information are inherently discrete. As a result, marginal costs are generally driven below average costs, which makes marginal cost pricing economically unviable. Furthermore, there is a tendency towards monopolization of such markets.

The third source of market failure in the production of information and knowledge stems from the *public-good* nature of these two goods. Producers of public goods generate many externalities and it is well known that in the presence of externalities, markets may not provide

[21] These numbers are reported by PhRMA, the trade organization of the US pharmaceutical firms, on its website: www.phrma.org

the right incentives to produce. Public goods are characterized by *non-rivalness* in consumption: the consumption of the good by one person does not prevent (rival) its consumption by another person; there is thus a potential for collective consumption. Another way to define non-rivalness is to say that a good is non-rival if for any given level of production, the marginal cost of providing it to an additional consumer is zero. This is clearly the case for knowledge and information. Think of a new idea or of a song performed in public. Public goods are said to be 'pure' when they are also *non-excludable*, in the sense that one person cannot exclude another person from consuming the good in question. Whereas non-rivalness is an attribute of the good itself, excludability depends, at least in part, on the available technology for exclusion and the institutional (legal) framework that permits or facilitates such technically feasible exclusion.[i]

The three sources of market failure associated with information production generate a *problem of appropriability*. That is, innovators and creators face a serious risk of appropriating only incompletely the returns from their activities. As a result, there is a general presumption that markets provide too little incentive to introduce new innovations and that the production of information and knowledge may well be insufficient from a social point of view. Economists therefore agree that governments ought to intervene to promote *dynamic efficiency*, that is, to provide the right *incentives* to create and innovate. We turn now to various responses, both public and private, to the problem of appropriability.

19.1.2 Intellectual property rights

IP rights refer to the legal rights which result from intellectual activity in the industrial, scientific, literary and artistic fields. These rights attach not only to inventions, business methods, industrial processes, chemical formulae, unique names (the so-called 'industrial property branch' of IP), but also to all information products that derive their intrinsic value from creative expression, literary creation, ideas or presentations (the so-called 'copyright branch' of IP). For instance, the text that you are currently reading, the articles and books that are referred to in this text, and the various pieces of software that have been used to produce this text are all intellectual properties.

Incentives versus use

Most countries have adopted laws to protect intellectual property. The main objective of IP law is to promote innovation and aesthetic creativity. To solve the appropriability problem, IP law intends to make knowledge or information excludable by legal means. That is, IP law grants exclusive use of the protected knowledge or creative work to the creator. Thereby, IP law provides creators with the incentives to produce new knowledge and partly solves the *underproduction* problem that would have resulted from the non-excludability of knowledge. However, by granting exclusive – monopoly – rights to the creator, IP law creates an *underutilization* problem. Indeed, as the

[i] To understand this difference, just look at television programme services in most countries. A television signal is clearly non-rival since, once it is broadcast, the marginal cost of making the broadcast available to another user is zero. Actually, to watch terrestrial TV broadcasts, all you need is TV receiving equipment. Yet, to use such equipment, you may be required by law to pay a special tax: those who fail to pay are legally – but not technologically – excluded from watching terrestrial broadcasts. On the contrary, cable TV broadcasts are made exclusive by encrypting the signal and charging for the device that allows it to be decoded. A similar distinction holds for knowledge: as we explain in this chapter, most countries have adopted laws to 'protect' intellectual property, that is, to allow producers of IP to prevent non-payers from using it.

marginal cost of production is zero (knowledge and information are non-rival goods), any positive price creates a welfare-reducing rationing.

In order to strike a balance between these two conflicting problems, IP law grants exclusive rights only for a limited period of time. That is, IP law addresses the two problems sequentially. First, legal protection makes the good excludable: to enjoy the services, users have to pay royalties to the producer. Second, once the protection is over, the good falls in the public domain, which means that all users may access the good for free (i.e., at marginal cost).

Lesson 19.1 **IP law attempts to find the best possible compromise between dynamic efficiency considerations (how to provide the right incentives to create and innovate) and static efficiency considerations (how to promote the diffusion and use of the results of creation and innovation).**

Note that dynamic efficiency calls for the broadest and longest possible protection (to maximize the flow of new knowledge creation), whereas static efficiency calls for the absence of protection (to avoid the deadweight loss of monopoly). Therefore, the balance struck by IP law is necessarily imperfect as it is impossible to reach both objectives at the same time.

Main IP regimes

The IP protection legislation generally distinguishes among four separate IP regimes, which are targeted at different subject matters. On the one hand, *patents*, *trade secrets* and *trademarks* are designed to protect industrial property (such as inventions, processes, machines, brand names, industrial designs, etc.). On the other hand, *copyrights* concern literary, musical, choreographic, dramatic and artistic works (such as novels, poems, plays, films, songs, drawings, paintings, photographs, sculptures, architectural designs, etc.).

In this chapter, we will focus mainly on patents and on copyrights.[j] Although both regimes aim to strike a balance between incentives and use, they accomplish this objective in very different manners. In short, patent law provides inventors with a strong and broad form of protection, but over a relatively short period of time. By contrast, copyright law affords a weaker and narrower form of protection (it merely protects expression, not the underlying ideas) but for a longer duration of time. Table 19.1 compares the two main IP regimes.[22]

The gradual reinforcement of IP protection

Under the initiative of the USA and of Europe, *IP protection has been strengthened, broadened and harmonized internationally*. In terms of *strengthening*, in the early 1980s, legal and procedural reforms in the USA provided stronger protection to holders of existing patents;[k] in Europe, the

[j] We describe trade secrets below. Trademark protection is designed to protect the integrity of the commercial marketplace, rather than to promote innovation.

[22] This table is adapted from Menell and Scotchmer (2007) and from Strowel (2005).

[k] For instance, the Patent and Trademark (Bayh–Dole) Act of 1980 allows universities and other non-profit organizations to patent discoveries made in their laboratories. Also, the Court of Appeals for the Federal Circuit was established in 1982 to harmonize patent law nationwide, which had the effect of strengthening patent protection.

Table 19.1 *Comparative overview of patent and copyright protection in the EU and in the USA*

	Patent	Copyright
Requirements for protection	(EU) Novelty, inventive step, industrial use (US) Novelty, non-obviousness, utility	Originality, authorship, form of expression
Ownership	(EU) First to file (US) First to invent	Author/creator
Rights	*Bundle of rights extending to the idea*: exclusive rights against all commercial uses (make, use, sell the innovation)	*Economic and moral rights on the form of expression*: exclusive rights against copying (rights of performance, display, reproduction, derivative works)
Scope of protection	Wide	Narrow
Duration	20 years from filing	Life of author + 70 years
Costs of protection	Filing, issue and maintenance fees; litigation costs	No filing necessary; suit requires registration; litigation costs

European Patent Office (EPO) granted the first European patents in 1978, but a genuine European patent (superseding national patents) is still under debate. Regarding *broadening* IP, new categories of invention have been protected, either through an extension of patent protection (software, business methods, genetic inventions) or through the creation of 'sui generis' rights (semiconductors, databases). Finally, the TRIPS Agreement of 1994, negotiated within the framework of the World Trade Organization, represents a major advance towards the *harmonization* of IP laws; it includes a general definition of patents, which adopts US criteria and, thereby, broadens the scope of patentable inventions internationally; furthermore, the USA and the EU repeatedly concluded bilateral agreements with their trading partners in order to coerce them to significantly strengthen their own IP rights regimes.

19.1.3 Alternative incentive mechanisms: rewards and secrecy

Rewards and secrecy are two alternative incentive mechanisms to promote innovation and aesthetic creativity, which do not entail as large a deadweight loss as IP rights, but which might create other problems.

Reward systems

Governments fund technical and artistic works through in-house development, procurement through competitive bidding, research grants to universities and promising scientists, or patronage of artists. Such funding mechanisms enhance static efficiency with respect to IP protection: as there is no need to grant exclusive rights to the innovator, the innovation is in free access and no deadweight loss ensues. However, to fund research and creation, governments have to raise taxes, which introduces distortions elsewhere in the economy and reduces static efficiency. In contrast, the patent system assigns costs to users rather than to tax payers. Moreover, there is no guarantee that subsidies achieve dynamic efficiency. Indeed, the uncertainty surrounding the social value of an innovation might yield the government over- or underestimating the amount of subsidy and, thereby, giving too much or too little incentive. In contrast, the patent system can be implemented without requiring sensible economic information that is only privately known.

To address the latter problem, several methods are used to ensure that a prize reflects the value of the innovation. For instance, by giving the inventor the option to choose IP protection instead, the prize effectively becomes a patent *buyout*: the inventor will enter the scheme only if the prize is at least as large as the value of the underlying patent, which constrains the prize to reflect the value. Another method consists of making the prize conditional on a verifiable performance standard, as illustrated in Case 19.1.

Case 19.1 The 'H-Prize'[23]

The H-prize was passed in May 2006 by the US House of Representatives. This national prize competition would help overcome technical challenges related to hydrogen. Modelled after the successful X Prize (which spurred the first privately funded suborbital human spaceflight in 2005), the H-prize offers prizes in three categories: (i) technological advancements (four prizes of up to \$1 million awarded biennially in the categories of hydrogen production, storage, distribution and utilization); (ii) prototypes (one prize of up to \$4 million awarded biennially that forces working hydrogen vehicle prototypes to meet ambitious performance goals); and (iii) transformational technologies (one grand prize consisting of a \$10 million cash award).

Trade secrets

In the absence of any external mechanism, inventors might sometimes find sufficient incentives to innovate when they manage to keep their discoveries secret. Famous examples are the Michelin radial tyres and the recipe for Coke, which have never been deconstructed or revealed. As long as the innovation is kept secret, information is excludable and the appropriability problem disappears. Firms may prefer to protect their discoveries through secrecy because they find that seeking a patent is, comparatively, a long and costly process. They might also wish to exploit their discoveries over a longer period than the duration of the patent. However, secrecy might be hard to keep as the risk is high that an employee (or some industrial spy) will disclose the invention, which would then become public knowledge. Trade secrets partly reduce such risk. Trade secret law protects the inventor against individuals within the laboratory or firm and those subject to contractual limitations who misappropriate proprietary information; to obtain this protection, the inventor needs merely to take reasonable steps to maintain secrecy. But, even if the costs of keeping secrets

[23] See H-Prize Act of 2006 (H.R. 5143).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

are reduced, the innovator loses all protection once the idea is released. Also, trade secrets offer no protection against independent innovations (which, by contrast, patents do).

Although secrecy makes information excludable, which solves the underproduction problem, secrecy does not solve the underutilization problem as information remains non-rival. Hence, the absence of diffusion creates a cost for society. This cost is reduced in the patent system for two reasons. First, as already explained, patent protection is limited in time: at the end of the protection, the innovation falls in the public domain. Second, patents entail a disclosure requirement: applicants must describe their invention in sufficient detail for a skilled person or team to be able to reproduce it. Knowledge is thus diffused, which fosters technical progress.

19.1.4 Protection of IP in practice

Besides the public responses designed to alleviate the appropriability problem of innovation, there also exist a number of private responses. As we have just seen, innovators might try to keep their discoveries secret. They might also take future imitation as a fact and simply rely on the (temporary) competitive advantage conferred by their innovative position. When interrogated, innovators often declare that they prefer such private measures to patents. However, the number of patents filed and granted has exploded over the last three decades. Is there a contradiction? These are the questions we address in this subsection.

Survey of innovators

Several empirical studies have attempted to assess the relative attractiveness of the different means innovators have at their disposal to protect their inventions. Interestingly, it appears that innovating firms consider trade secrets (for process innovations) and business strategies based on early-mover advantage (for product innovations) as the main means of getting returns on R&D investments and to appropriate the rents stemming from innovation. Similarly, according to a recent survey, managers claim that ‘lead time, learning curves, and sales or service efforts are substantially more effective in protecting IP than patents are’.[24]

It appears thus that the appropriability problem is often better addressed through private responses than through public responses. In particular, except for the chemical and pharmaceutical sectors (see Case 19.2 below), patent protection is generally deemed to be of little efficacy, especially for process innovations. We can give a number of reasons for this lack of efficacy: (i) that a patent can easily be ‘invented around’ by imitators; (ii) that a patent is costly to obtain and to enforce; and (iii) that innovators suffer from disclosing the information, as required by the patent.[25]

Case 19.2 Patents in the pharmaceutical sector

The three market failures we identified above are particularly acute in the pharmaceutical sector: large indivisibilities result from huge R&D fixed costs (the average cost of a new molecule is estimated at EUR 1.25 billion); the length of the R&D process and the need to get

[24] See Anand and Galetovic (2004), which is based on a survey of 600 managers.
[25] See the survey of recent studies by Caillaud (2003).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

public approval for new drugs cause a lot of uncertainty (of 10 000 pharmaceutical products patented, only 10 are marketed);[26] finally, because knowledge is more science-based and more codified in the pharmaceutical sector, imitation costs are low and hence, externalities are important. It must be added that successful pharmaceutical innovations create other powerful consumption externalities as they improve public health. For all these reasons, the discrepancy between social and private returns to innovation is particularly wide in the case of pharmaceuticals and thus, absent appropriate intervention, the level of pharmaceutical innovation would undoubtedly be insufficient from a social point of view.

Does the patent-based system constitute the appropriate public intervention for the pharmaceutical sector? In terms of *dynamic efficiency*, empirical studies indicate two reasons for concluding that the patent system is relatively more efficient in the pharmaceutical sector than in other industrial sectors. First, it appears that pharmaceutical companies rely heavily on patents to appropriate the returns from their inventions.[27] Second, macroeconomic analyses also show that the pharmaceutical sector plays a leading role when looking at the patent explosion (especially in the USA). Note that patents are relatively efficient in the pharmaceutical sector despite the fact that the effective patent life (i.e., the patent time remaining at the product launch) is significantly reduced for drugs. Indeed, patents in pharmaceuticals are typically applied for early in the development process; because of the length of the regulatory approval process, marketing exclusivity occurs only after a number of years. In consequence, the USA, Europe and Japan have enacted patent term restoration laws.

On the contrary, the *static inefficiency* resulting from monopoly prices is worse for pharmaceuticals than for most other products. To put it bluntly, the rather abstract 'deadweight-loss of monopoly' takes here a much more concrete and tragic form, which can be measured, following the World Health Organization (WHO), in *disability-adjusted life years* (DALYs) lost.[1] Therefore, one understands why, in many developed countries, the prices of drugs are controlled and/or expenditures for drugs are covered by public or private insurance. While such interventions alleviate the static inefficiency of patents, they also distort market incentives, as potential demand levels condition research efforts. The latter consideration stresses another shortcoming of the patent-based system: as R&D priorities are decided on the basis of potential demand levels, the system provides few incentives to develop products with relatively small economic markets. What makes the matter worse is that those products are generally of great social need.[28] It thus appears that in the pharmaceutical sector, the patent-based system needs to be complemented by adequate subsidy-based approaches.

[26] See OECD (2000).
[27] See Mansfield (1986), Levin *et al*. (1987), Cohen, Nelson and Walsh (2000) and Arora Ceccagnoli and Cohen (2003).
[1] According to the WHO, 'DALYs for a disease are the sum of the years of life lost due to premature mortality in the population and the years lost due to disability for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death to include equivalent years of "healthy" life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health'.
[28] As Ridley, Grabowski and Moe (2006) explain it, '[t]hree diseases with the greatest burden are HIV/AIDS, malaria, and tuberculosis. Manufacturers do invest in R&D for HIV/AIDS, because there is a market in both developed and developing countries for these therapies. There is less incentive to invest in R&D for malaria, because more than 99 percent of disability-adjusted life years (DALYs) lost to malaria are in developing countries'.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

The patent explosion

In spite of what innovators declare about the relative unimportance of patents to protect their IP, the number of patent applications and grants has risen drastically over the last decades. In the USA, it has more than tripled between 1980 and 2001 (whereas it was practically stable over the previous two decades). A comparable trend is observed for European countries (although it began later). Although nearly all technology fields experienced growth in patenting, two technology fields contributed substantially to the overall surge in patenting: biotechnology and information and communication technologies.

However, this growth in the number of patents does not necessarily mean that the total value of innovations follows the same increasing path. Ultimately, what we would like to measure is the consumer value created by R&D spending. As long as R&D spending and patents are linked, the issue is thus to measure the private value of patents. Yet, this issue is far from simple as the private value of receiving a patent depends on the counterfactual. That is, what would happen if no patent was granted? There are four possible scenarios: either (i) the invention is not made at all, or the invention is made but (ii) the patent is granted to a rival firm, or (iii) the invention is put in the public domain, or (iv) the invention is kept secret and not patented. Scotchmer (2004, p. 275) summarizes the main results drawn from various estimations of patent values: '(1) the values of patent rights are very dispersed, (2) the distribution of values is very skewed, with most of the value provided by a few high-earning patents, and (3) the average value of patent rights is much lower than the average R&D cost of innovation'.

A patent paradox?

The huge increase in patent counts seems to contradict what innovators declare about the relative unimportance of patents to protect their IP. How can we solve this so-called 'patent paradox', which has been systematically documented in the semiconductor industry?[29] For a number of reasons, firms in most sectors still bother to seek patent protection for their inventions:[30] (i) patents are relatively inexpensive to register (although they are generally costly to defend); (ii) patents can serve to measure the ouput of a firm's R&D division and, thereby, to structure compensation and incentive schemes; (iii) venture capitalists often demand that firms patent technology, both to block rivals and to have assets to sell in case the firm flounders; (iv) patents can be used as a 'trading device'.

The last reason is confirmed by a number of surveys, which show that it is essentially large firms that resort to patent protection, and especially in complex industries (e.g., biotech, IT, telecoms, electronics and software).[m] We will return to this *patent portfolio theory* in Section 19.3. The main idea is that, in many industries, patents are more valuable when aggregated than when

[29] Hall and Ziedonis (2001) interviewed industry representatives and observed (over the period 1979–95) that (i) firms did not rely heavily on patents to appropriate returns to R&D, but (ii) their propensity to patent rose dramatically after the mid-1980s.

[30] See Geroski (1995).

[m] A survey published in *The Economist* (20 October 2005) reports the following facts: IBM now earns over $1 billion annually from its IP portfolio; HP's revenue from licensing has quadrupled in less than three years, to over $200m this year; Microsoft is on course to file 3000 patents this year, when in 1990 it received a mere 5; 54% of companies saw growth in licensing of 10–50% between 2000 and 2002; almost 75% of executives say they expect to buy as well as sell more licences over the next two to five years, and 43% expect a dramatic increase in their licensing revenue (according to a survey by McKinsey).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

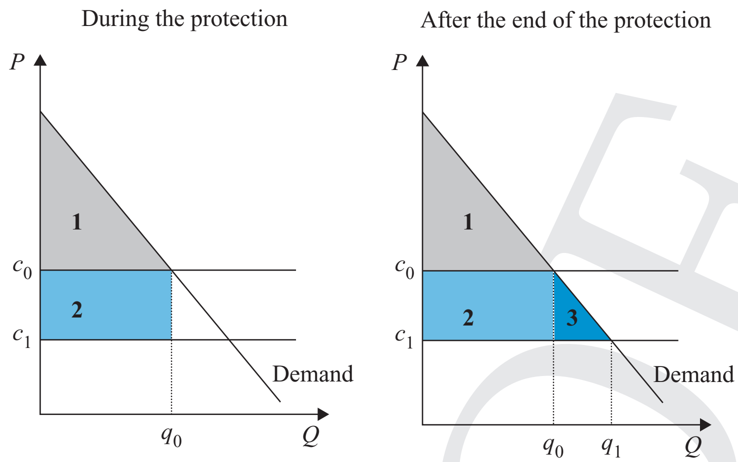

Figure 19.1 Trade-off between dynamic and static efficiency

taken individually. Firms have thus an incentive to constitute large 'portfolios' of related patents. This is particularly true in complex industries that rely heavily on *cumulative innovations*.

19.2 Protecting innovations

The two objectives of protecting innovations are, on the one hand, promoting the diffusion and use of new innovations while, on the other hand, preserving the incentives to innovate in the first place. In this section, we examine how three different institutions, namely IP rights, rewards and trade secrets, balance these static and dynamic efficiency considerations.

19.2.1 Optimal design of IP rights

We study here how to design IP rights so as to achieve the best possible balance between incentives and diffusion. We first give a more precise description of the trade-off between static and dynamic efficiency; we examine then, more formally, the optimal choice of two dimensions of IP rights, namely length and breadth.

Static vs. dynamic efficiency

To understand better the trade-off between static and dynamic efficiency, we use the simple model illustrated in Figure 19.1. Consider an industry where an arbitrary number of firms produce a homogeneous good at constant marginal cost c_0. Firms are assumed to compete à la Bertrand. Therefore, at this stage, the price is equal to c_0, a quantity q_0 is sold, the consumer surplus is represented by area 1 and the producer surplus is nil. Suppose now that some firm i discovers a process innovation that reduces the marginal cost of production from c_0 to c_1. (For simplicity, we do not model explicitly the R&D cost that is associated with discovering the innovation.) We assume that the innovation is non-drastic (i.e., as defined in Chapter 18, the cost advantage generated by the innovation does not allow the innovator to escape totally from the competition of the other firms in the industry).

Absent IP protection, as knowledge is non-rival and non-excludable, all firms would possibly have access to the innovation.[n] As a result, firm i would anticipate that the innovation would not generate any extra profit and would therefore refrain from investing in R&D. This would be detrimental for society, as a more efficient production technology would not be used. By granting exclusive rights on the innovation, IP law provides firm i with incentives to innovate. Firm i then becomes the only firm producing at marginal cost c_1. Because the innovation is supposed to be non-drastic, the monopoly price corresponding to c_1 is larger than the marginal cost of the rival firms (c_0). Therefore, as we saw in Chapter 3, the equilibrium is such that all firms set a price equal to c_0 and all consumers buy from the innovator, who then sells a quantity q_0.[o] The consumer surplus is still given by area 1 and social welfare is now augmented by area 2, which represents firm i's profit.

However, as long as the innovation is protected, social welfare does not reach its maximum. Indeed, from the point of view of static efficiency, the quantity sold by firm i falls short of the optimal quantity corresponding to cost c_1 (i.e., q_1). It is only when the legal protection ends that this quantity is produced: all firms have access to the new technology and the market price falls to c_1. The producer surplus is nil again but the consumer surplus now amounts to areas 1, 2 and 3. In sum, area 3 can be seen as the temporary deadweight loss society has to pay to make sure that the innovation takes place: *a temporary reduction in static efficiency enhances dynamic efficiency*.

The next question that arises naturally is: how long should this reduction in static efficiency last? As far as patents are concerned, the 1994 WTO's *Agreement on Trade-Related Aspects of Intellectual Property Rights* (*TRIPS*) has harmonized the answer to this question: patent protection must be available for at least 20 years. Why 20 years? Is such a duration appropriate? To address this issue, we turn now to a more elaborate model.

Optimal patent length

We consider an innovator with a strictly convex cost function[31]

$$C(x) = \tfrac{1}{2}\phi x^2,$$

where ϕ reflects the exogenous efficiency of the existing innovation technology (the lower ϕ, the more efficient the innovation technology). We assume that ϕ is large enough, so that in all circumstances $x \leq 1$ and, accordingly, x can be seen as the success probability of the innovation. That is, with probability x, the innovation is succesful and the innovator obtains the monopoly profit π^m during the life of the patent, and some competitive return $\bar{\pi}$, with $0 \leq \bar{\pi} < \pi^m$, once the

[n] This implicitly assumes that the innovator does not take any effort to conceal information (see the discussion of secrecy below).

[o] This assumes that in case of equal prices, consumers buy from the most efficient firm. Although there is a continuum of symmetric equilibria where all firms set a price p lying between c_1 and c_0, the one we select here is the only equilibrium that resists small perturbations (see the discussion in Chapter 3). Alternatively, we could say that the innovator sets a price one cent below c_0 and attracts all consumers because it is less expensive than the other firms.

[31] The first rigorous model explaining the fundamental trade-off between static and dynamic efficiency was proposed by Nordhaus (1969). Since this seminal paper, there has been extensive research on the consequences of patent protection for social welfare. We summarize here the main finding of this research in a simplified framework. The exposition follows Takalo (2001).

patent has expired and the innovation has become available to everyone.[p] Noting the patent length T, we can express the present discounted value (in continuous time) of the innovator's return in case of success as

$$P(T) = \int_0^T e^{-rt}\pi^m dt + \int_T^\infty e^{-rt}\bar{\pi}\, dt.$$

The innovator's problem is thus to choose x so as to maximize $xP - (1/2)\phi x^2$. The solution is easily found as

$$x^*(T) = P(T)/\phi.$$

As $\bar{\pi} < \pi^m$, $P(T)$ is clearly an increasing function of T, and so is $x^*(T)$: the longer the patent protection, the higher the innovator's return and, hence, his incentive to invest in R&D.

Similarly, we can express the social return on innovative effort as

$$S(T) = \int_0^T e^{-rt}W^m dt + \int_T^\infty e^{-rt}\bar{W} dt,$$

where W^m and $\bar{W}$ depict social welfare (computed as the sum of consumer surplus and industry profits) when the patent is in force and after it expires.[q] Contrary to $P(T)$, $S(T)$ decreases with T (since $W^m < \bar{W}$), which sets the crucial distinction between the social and private return on innovation. The policymaker's task is to choose the patent length that maximizes $S(T)$, given that it correctly anticipates the innovator's profit-maximizing research intensity $x^*(T)$. That is, the policymaker's objective can be written as

$$\max_T x^*(T)\, S(T) - \tfrac{1}{2}\phi\left(x^*(T)\right)^2.$$

Slightly rearranging terms, we can rewrite the first-order condition as

$$\underbrace{\frac{\partial x^*(T)}{\partial T} S(T)}_{\text{marginal dynamic gain}} = \underbrace{x^*(T)\left(\phi \frac{\partial x^*(T)}{\partial T} - \frac{\partial S(T)}{\partial T}\right)}_{\text{marginal static loss}}. \tag{19.1}$$

The latter condition illustrates the trade-off between the static and dynamic efficiency considerations facing the policymakers. The optimal patent length should be chosen so that the marginal dynamic gain of prolonged protection is equal to the marginal static loss. The marginal dynamic gain is measured by the left-hand side of (19.1), which shows how an increase in patent life encourages innovative endeavours. The marginal static loss is measured by the right-hand side of (19.1); it comprises two terms: the increased R&D cost due to the accelerated innovative effort and the decrease in consumer surplus resulting from a longer innovator's monopoly.

The main conclusion from this model is that *the optimal patent duration is finite*. There are indeed two forces that work to limit the optimal length of a patent. First, there are diminishing returns to R&D activity: because the cost function is convex, it becomes progressively more expensive to increase the probability of success and therefore, it will take progressively greater increases in T to achieve a given probability of success. The second force is discounting: the consumer benefits from the innovation will not be realized until after the patent expires and so, the larger T, the smaller the present value of those benefits.

[p] In the model of Figure 19.1, π^m corresponds to area 2, while $\bar{\pi} = 0$.
[q] In the model of Figure 19.1, W^m corresponds to the addition of areas 1 and 2, while $\bar{W}$ is W^m augmented by area 3.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

Lesson 19.2 **A patent that is unlimited in duration cannot be welfare-maximizing.**

The argument carries over to copyright protection. As illustrated in Case 19.3, it is exactly the framework we have just presented that was used by a number of economists to express their disagreement with the extension of the copyright term in the USA.

Case 19.3 Arguments against the extension of copyright term in the USA

Voted in 1998, the Copyright Term Extension Act (CTEA, aka Sonny Bono Copyright Act) extended the duration of existing US copyrights by 20 years (i.e., to life of the author plus 70 years and for works of corporate authorship to 120 years after creation or 95 years after publication, whichever endpoint is earlier). In 1999, a group of commercial and non-commercial interests who relied on the public domain for their work (led by Eric Eldred, an Internet publisher) challenged the constitutionality of the CTEA. In 2002, 17 economists (among them five Nobel laureates) supported the petitioners by submitting an *amicus curiae* brief (i.e., some information voluntarily offered by a 'friend of the court' to assist the court in deciding a matter before it). The first section of the brief was entitled: 'It is highly unlikely that the economic benefits from copyright extension under the CTEA outweigh the costs.' To justify this statement, the economists used the above framework: they argued that the revenues earned during the additional 20 years of protection are so heavily discounted that they lose almost all value, while the extended protection of existing works generates immediate deadweight losses (which are even larger when taking the increased cost of creating new derivative work into account). Despite this support, the Supreme Court found against the petitioners (see Eldred *et al. vs.* Ashcroft, Attorney General, 537 U.S. 186, 2003).[32]

Optimal patent breadth

Fixing a finite patent duration is not the only way through which policymakers can avoid excessive monopoly power. The extent of monopoly power can also be curbed by limiting the *breadth* of the patent. The meaning of patent breadth is relatively vague.[r] The breadth measures the *degree of patent protection*, but several measures have been proposed. Basically, economists study breadth in two types of innovative environment: where the innovation is threatened by horizontal competition, or where an innovation might be supplanted by an improved innovation. We consider here the former environment; we will consider the latter when we examine cumulative innovations.

As far as horizontal competition between innovations is concerned, economists model breadth in two ways: either in the 'product space', by defining how similar a product must be to

[32] See Liebowitz and Margolis (2004) for a deeper analysis of this case.

[r] The notion of patent breadth is not directly defined in the IP law. Breadth is actually a matter of interpretation. On the one hand, the patent office will evaluate the patentability of the innovation (is the innovation – using the European terminology – novel, inventive and industrially applicable?) and the legitimacy of the claims put forth by the patentee along with the description of the innovation. On the other hand, courts will judge whether there is infringement.

infringe a patent; or in the 'technology space', by defining how costly it is to find a non-infringing substitute for the protected market. In the first interpretation, a broader patent excludes a larger set of horizontal substitutes; in the second interpretation, a broader patent makes it more costly to enter the market with an alternative technology (i.e., to invent around the patent). We simplify matters here by assuming that the innovator's profit and social welfare are functions of patent breadth, which we measure by the parameter $b \in [0, 1]$. Accordingly, we note

$$\begin{cases} \pi(b) \text{ with } \pi(1) = \pi^m \text{ and } \pi(0) = \bar{\pi}, \\ W(b) \text{ with } W(1) = W^m \text{ and } \pi(0) = \bar{W}. \end{cases}$$

As $\pi'(b) > 0$ and $W'(b) < 0$, patent breadth exerts, like patent length, opposite effects on the innovator's profit and on social welfare. The private and social returns on innovation can now be rewritten as

$$P(T, b) = \int_0^T e^{-rt}\pi(b)\,dt + \int_T^\infty e^{-rt}\bar{\pi}\,dt,$$

$$S(T, b) = \int_0^T e^{-rt}W(b)\,dt + \int_T^\infty e^{-rt}\bar{W}\,dt.$$

Regarding the innovator, we just need to re-express the first-order condition as

$$x^*(T, b) = P(T, b)/\phi. \tag{19.2}$$

Totally differentiating (19.2) yields

$$\frac{dT}{db} = -\frac{\partial P/\partial b}{\partial P/\partial T} < 0, \tag{19.3}$$

which implies that *length and breadth are substitutable policy tools* with regard to innovation.

Lesson 19.3 **Public policy can use patent length and patent breadth as substitutes.**

The policymaker's task is now to find the optimal patent breadth–length mix, anticipating the innovator's optimal conduct. This problem is usually understood as the maximization of social welfare from existing innovation, constraining the supply of innovation to a predetermined level. Formally, it amounts to maximizing S with respect to T and b, fixing the innovation activity x at some required level. Doing so, we can define $T(b)$ by solving Equation (19.2) for T and express the social value of an existing innovation as a function of breadth only: $S(T(b), b)$. Differentiating the latter function with respect to b and using (19.3) gives

$$\frac{dS}{db} = \frac{\partial S}{\partial T}\frac{dT}{db} + \frac{\partial S}{\partial b} = -\frac{\partial S}{\partial T}\frac{\partial P/\partial b}{\partial P/\partial T} + \frac{\partial S}{\partial b}. \tag{19.4}$$

The optimal patent policy is determined by the sign of dS/db. There are two possible cases.

1. If $dS/db > 0$, then increasing breadth is welfare-enhancing. Consequently, *the optimal patent is broad and short*: it has maximum breadth and minimum length, that is, $b = 1$ and $T = \underline{T}$ (where $\underline{T}$ is defined as the value of T solving Equation (19.2)

for $b = 1$). Using Equation (19.4) and recalling that $\partial S/\partial T$ is negative, we have that

$$\frac{dS}{db} > 0 \iff \frac{\partial P/\partial b}{\partial P/\partial T} > \frac{\partial S/\partial b}{\partial S/\partial T}, \tag{19.5}$$

which can be interpreted by saying that an increase in patent breadth stimulates investment in innovation relatively more than patent length while reducing the post-innovation welfare relatively less. It is therefore optimal to set as short a patent life as possible and to extend patent breadth correspondingly.

2. If $dS/db < 0$, then increasing breadth is welfare-detrimental. Consequently, *the optimal patent is narrow and long*: it has minimum breadth and maximum length, that is, $b = \underline{b}$ and $T = \infty$ (where $\underline{b}$ is defined as the value of b solving Equation (19.2) for $T \to \infty$).[s] In this case, reversing the sign of inequality (19.5), we have that an increase in patent breadth curbs post-innovation social welfare relatively more and accelerates innovative activity relatively less than an increase in patent life; hence, it is desirable to make a patent as narrow as possible by prolonging patent life correspondingly; this leaves the incentive to innovate unaltered but expands static social welfare.

Lesson 19.4 **If the marginal rate of substitution of patent length for breadth is larger on the incentive to innovate than on social welfare, the optimal patent is broad and short; otherwise, it is narrow and long.**[33]

19.2.2 Rewards vs. patents

To compare the relative merits of rewards and patents as mechanisms to encourage innovation and diffusion, we build on the model that we used in Subsection 19.2.1 to examine the optimal patent length.[34] Recall that we consider an innovator with a strictly convex cost function, $C(x) = \frac{1}{2}\phi x^2$, where x can be seen as the success probability of the innovation, and ϕ reflects the exogenous efficiency of the existing innovation technology. Recall also our previous notation: when the innovator has the exclusive use of its innovation, per-period profits and welfare are equal to π^m and W^m whereas, if the innovation is in the public domain, they are equal to $\bar{\pi}$ and $\bar{W}$, with $\bar{\pi} < \pi^m$ and $\bar{W} > W^m$. Taking an interest rate r and assuming continuous time, we can compute the present discounted value of profits and social welfare for a given patent length T as

$$\begin{aligned} P(T) &= \int_0^T e^{-rt}\pi^m dt + \int_T^\infty e^{-rt}\bar{\pi}\, dt = \tfrac{1}{r}\bar{\pi} - \tau(\bar{\pi} - \pi^m), \\ S(T) &= \int_0^T e^{-rt} W^m dt + \int_T^\infty e^{-rt}\bar{W} dt = \tfrac{1}{r}\bar{W} - \tau(\bar{W} - W^m), \end{aligned} \tag{19.6}$$

[s] In the interests of analysis, it is assumed that the minimum values $\underline{T}$ and $\underline{b}$ exist, and are positive and finite.

[33] Using different models of R&D competition and different measures of patent breadth, a number of economists have reached opposite conclusions about the socially optimal patent length–breadth mix. Tandon (1982) and Gilbert and Shapiro (1990) call for narrow and long patents. In contrast, Gallini (1992) advocates broad and short patents, while for Nordhaus (1972), all mixes are equivalent. Klemperer (1990) provides examples of all these results. Denicolo (1996) and Takalo (2001) show that the above-mentioned rule can be applied to reconcile the seemingly contradictory policy implications drawn from a number of previous analyses.

[34] We follow here the analysis of Shavell and van Ypersele (2001).

where $\tau \equiv (1 - e^{-rT})/r$ is the discounting-adjusted length of the patent. Note that in the absence of patent protection (i.e., if the innovation passes immediately into the public domain), we have $T = \tau = 0$, so that $P(0) = \bar{\pi}/r$ and $S(0) = \bar{W}/r$.

As a benchmark, we first compute the expected social welfare that would be achieved in a first-best world, where the public authority chooses R&D effort and puts the innovation immediately into the public domain. If the innovation is successful, the present-discounted value for society is $S(0) = \bar{W}/r$. The public authority's maximization problem is thus

$$\max_x x(\bar{W}/r) - \tfrac{1}{2}\phi x^2, \tag{19.7}$$

and the optimal R&D effort is easily found as (with the subscript FB for 'first-best') $x_{FB} = \bar{W}/(\phi r)$; the expected social welfare is then $W_{FB} = \bar{W}^2/(2\phi r^2)$.

Consider now a second-best world where it is the innovator who chooses the R&D effort on the basis of the private value of the innovation. As far as this private value is concerned, we want to compare two incentive mechanisms: patents and rewards. In the *patent system*, as we explained above, the private value of the innovation is equal to $P(T)$ and the innovator's optimal effort is $x_P = P(T)/\phi$ (with the subscript P for 'patent'). We can then compute the expected social welfare under the patent system as

$$W_P = x_P S(T) - \tfrac{1}{2}\phi x_P^2.$$

We note that *the patent system fails to achieve the first-best outcome*: $W_P < W_{FB}$. This merely confirms our previous analysis as we already argued that the patent system entails two sources of welfare loss relative to the first-best. First, there is a deadweight loss because of monopoly pricing: as $S(T) < S(0)$, we have that $W_P < x_P S(0) - \frac{1}{2}\phi x_P^2$; second, the investment in R&D is insufficient because the innovator can only appropriate the private (and not the social) value of the innovation: as x_{FB} maximizes $xS(0) - \frac{1}{2}\phi x^2$, we also have that $x_P S(0) - \frac{1}{2}\phi x_P^2 < x_{FB} S(0) - \frac{1}{2}\phi x_{FB}^2 = W_{FB}$.

Under a *reward regime*, the private value of the innovation is some reward R that the public authority grants to the innovator if he succeeds; the quid pro quo for this reward is that the innovation passes immediately into the public domain. The innovator's problem is thus given by

$$\max_x xR - \tfrac{1}{2}\phi x^2. \tag{19.8}$$

Comparing expressions (19.7) and (19.8), it is seen at once that it is possible to achieve the first-best outcome under the reward regime by simply setting $R = \bar{W}/r$, that is, by rewarding the innovator with the social value of the innovation. However, a prerequisite for this result is that the public authority possesses perfect information about the social value of the innovation, which is arguably quite unlikely (for instance, the innovator may have private information about the demand on the product market). Let us assume therefore that the public authority bases the reward on some estimation of the social value of the innovation, namely $R = (1 - s) S(0)$, with $-1 < s < 1$; that is, if $s > 0$ (resp. $s < 0$), the public authority underestimates (resp. overestimates) the social value of the innovation. In that case, the innovator optimally chooses an R&D effort equal to (with the subscript R for 'reward') $x_R = R/\phi = (1 - s) S(0)/\phi$. The resulting expected social welfare is then computed as

$$W_R = \frac{1}{2\phi}(1-s^2)S(0)^2 = (1-s^2)W_{FB},$$

which is clearly smaller than the first-best level W_{FB}. Note that the only distortion comes here from an inadequate investment in R&D (which may be too low or too high according to whether the social value of the innovation is under- or overestimated); as the innovation is immediately accessible to anyone, there is no deadweight loss due to monopoly pricing.

We can now compare the patent and the reward regimes. The patent regime should be preferred if $W_P > W_R$, which is equivalent to $W_P > (1-s^2)W_{FB}$, or

$$|s| > \sqrt{1-(W_P/W_{FB})},$$

that is, if the estimation of the social value of the innovation is sufficiently inaccurate.

Lesson 19.5 **Both the patent and the reward systems fail to achieve the first-best outcome. In the reward system, the investment in R&D is inadequate because the public authority cannot perfectly evaluate the social value of the innovation and fails thus to grant a reward that would achieve the first-best outcome. In the patent system, there is a deadweight loss due to monopoly pricing on top of an insufficient investment in R&D. Either system may be superior to the other; the patent system outperforms the reward system whenever the public authority fails to evaluate the social value of the innovation with enough accuracy.**

19.2.3 Secrecy vs. patents

The innovator may also try to capture the private value of the innovation by keeping the innovation secret. The risk of this strategy is naturally that the secret may eventually leak out, putting de facto the innovation into the public domain. We model this possibility by assuming that a competitor may discover the innovation according to a Poisson process with exogenous parameter $\lambda > 0$. This means that the probability that the innovation remains secret at any date t is equal to $e^{-\lambda t}$ (where the parameter λ measures the ease with which the competitor can discover or circumvent the trade secret innovation). We can then compute the present discounted value of profits and social welfare for a given λ as follows:

$$\begin{aligned} P(\lambda) &= \int_0^\infty (e^{-\lambda t}\pi^m + (1-e^{-\lambda t})\bar{\pi})e^{-rt}dt = \frac{1}{r}\bar{\pi} - \frac{1}{r+\lambda}(\bar{\pi}-\pi^m), \\ S(\lambda) &= \int_0^\infty (e^{-\lambda t}W^m + (1-e^{-\lambda t})\bar{W})e^{-rt}dt = \frac{1}{r}\bar{W} - \frac{1}{r+\lambda}(\bar{W}-W^m). \end{aligned} \tag{19.9}$$

Comparing expressions (19.6) and (19.9), and recalling that $\bar{\pi} < \pi^m$ and $\bar{W} > W^m$, we find easily that

$$P(T) > P(\lambda) \Leftrightarrow \tau > \frac{1}{r+\lambda} \quad \text{and} \quad S(T) > S(\lambda) \Leftrightarrow \tau < \frac{1}{r+\lambda},$$

where $1/(r+\lambda)$ can be seen as the discounting-adjusted duration of the secret. We can therefore record the following result.

Lesson 19.6 **Ignoring dynamic efficiency considerations, the interests of the innovator and of society are completely at odds when it comes to choosing between patents and trade secrets: the innovator prefers patents over secrets when patents have a longer duration (adjusted for discounting) than secrets, whereas the exact opposite prevails for society.**

To understand this result, note that the discounting-adjusted duration of the secret is inversely related to the 'probability of discovery' λ. Hence, the larger λ, the more risky it is to keep a trade secret and the more attractive it is to patent the innovation. In contrast, the opposite goes for society as the faster secrets are discovered, the faster social welfare increases from W^m to $\bar{W}$.

If we now take the innovator's choice of R&D effort into account, we quickly see that these two conflicting forces prevent us from drawing any clear conclusion. If the innovator opts for secrecy, he will invest $x_S = P(\lambda)/\phi$ and the expected social welfare will be given by $W_S = x_S(S(\lambda) - \frac{1}{2}\phi x_S)$. With a patent, we computed above that $W_P = x_P(S(T) - \frac{1}{2}\phi x_P)$. The only result we can establish with certainty is that $W_S = W_P$ if $\tau = 1/(r+\lambda)$. Yet, for $\tau > 1/(r+\lambda)$, we have on the one hand that $x_P > x_S$ but on the other hand that $(S(T) - \frac{1}{2}\phi x_P) < (S(\lambda) - \frac{1}{2}\phi x_S)$. A similar ambiguity prevails for $\tau > 1/(r+\lambda)$.

19.3 Cumulative innovations

Innovations are rarely produced in complete isolation. In many complex industries (such as biotechnologies, information technologies, telecommunications, electronics and software), innovations are typically the result of a *cumulative process*. For the sake of the analysis, it is useful to distinguish between two types of cumulativeness. On the one hand, innovations often build on previous inventions; a famous quote attributed to Isaac Newton nicely illustrates this idea: 'If I have seen further it is by standing on the shoulders of giants.'[35] We use the term *sequential innovations* to refer to the fact that a particular innovation leads to second-generation innovations. For instance, the invention of the laser led to surgical applications, spectroscopy, CDs and DVDs, etc. On the other hand, it is often the case that separate inventions have to be combined to create value. One talks then of *complementary innovations*: a second-generation product requires the input of a number of different first-generation innovations. Think of firms in the electronics industry (e.g., trying to produce new peripherals to be coupled with personal computers or video game consoles) or in the biotech industry (e.g., combining patented genes to bioengineer a new crop seed).

In this section, we explain why cumulative innovations raise a number of problems that the patent system, without further institutional provisions, is ill-equipped to cope with. First, the main problem with sequential innovations is that a patent on the first-generation innovation confers the patentee a *holdup* right over subsequent innovations; we will see how *ex ante licensing* might alleviate this problem, while the opportunistic conduct of some firms might exacerbate it. Second, the main problem for complementary innovations, referred to as the 'tragedy of the anticommons', is that the allocation of IP rights to separate right-holders results in higher total prices (a phenomenon also known as 'royalty stacking'). We will see in this case that solutions may come from *cross-licensing* and *patent pools*. Before that, we discuss the case of the smartphone industry, where the problems of cumulative innovations are particularly acute.

[35] Letter to Robert Hooke, 15 February 1676.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

Case 19.4 The smartphone patent wars

The global mobile phone industry is big, with a value of US$1.18 trillion in 2011. In April 2012, Lucy Koh, a US District Court Judge, tried to force a truce on one of the fiercest fronts of what has been dubbed the 'Smartphone Patent Wars'. She ordered Tim Cook and Choi Gee-Sung, the CEOs of Apple and Samsung, to meet and work out their bitter and prolonged patent disputes. The two firms had indeed spent long months suing and countersuing each other, all over the world, for alleged patent infringements. As most companies in the smartphone industry are fighting similar battles, the peacemaking process that Lucy Koh initiated may prove crucial for saving the industry from a potential collective suicide. The legal quagmire that plagues the smartphone industry results from a combination of forces. First, smartphones are the archetypal example of a cumulative innovation as they exhibit the two types of cumulativeness. There is complementarity because smartphones result from the collision between computers and mobile phones, which are themselves already cumulative in nature. There is also sequentiality because smartphones have to incorporate a number of existing standards (e.g., Wi-Fi access, email transfers or video display).

Second, the problems raised by complementarity and sequentiality of innovations are magnified by the sheer density and size of the 'patent thicket' that engulfs the smartphone industry: according to estimates, a smartphone is open to a quarter of a million patent claims. Factoring in the diversity of players (as companies flock from industries as diverse as equipment manufacturing and software development) and the crucial need for interoperability (just try to imagine how painful your life would be if several incompatible email services were coexisting), it is easy to understand why even though private solutions to the previous problems do exist, these solutions are very hard to achieve in the smartphone arena.

From a social point of view, these unceasing patent disputes are not only a waste of resources but, more fundamentally, they may stifle innovation in the industry. As Shapiro (2001, p. 126) nicely puts it, 'In these industries, the danger that a manufacturer will step on a land mine is all too real. The result will be that some companies avoid the mine field altogether, that is, refrain from introducing certain products for fear of holdup. Other companies will lose their corporate legs, that is, will be forced to pay royalties on patents that they could easily have invented around at an earlier stage, had they merely been aware that such a patent either existed or was pending. Of course, ultimately the expected value of these royalties must be reflected in the price of final goods.'

19.3.1 Sequential innovations and holdup

Patents have a double purpose: to protect R&D investments *and* facilitate the diffusion of knowledge (other researchers will benefit from the publication of the patent; future research might also be improved indirectly as open questions can be better identified). Obviously, the latter purpose becomes even more important in the presence of sequential innovations. This forces us to reconsider the previous analysis of the optimal patent length–breadth mix and address two additional important questions. (i) Should the initial innovator have a right on subsequent innovations? (ii) Insofar as subsequent innovations are not necessarily substitute to the initial innovation (they use it and reproduce it), should they be considered as infringements?

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

Intuitively, in order to provide each innovator with the proper incentives, we would like to answer 'yes' to the first question and 'no' to the second. Indeed, as earlier innovators provide benefits to later innovators, they should be compensated for their contributions by being granted some right on subsequent innovations. Also, if all subsequent innovations were considered as infringements, later innovators would have no incentive to invest. Yet, the larger the rights granted to the initial innovator, the lower the incentives for subsequent innovators, and vice versa. The problem is thus to determine how profit should be divided between successive innovators. This is a delicate problem and, unfortunately, IP appears to be a blunt instrument in that respect.

More precisely, sequential innovations may give rise to the classic economic problem of *holdup*: if an early patent holder has a claim against subsequent innovators, the latter may be reluctant to invest in R&D because they anticipate the expected cost of such claims. In what follows, we first examine the holdup problem in a simple model; we discuss next the case of new types of IP intermediary, which have developed a business model around a form of holdup.

A simple model of holdup

To illustrate the potential holdup problem with sequential innovations, we use the following simple model.[36] Suppose there are two sequential innovations, $i = 1, 2$; innovation i has (private and social) value v_i and can be produced by firm i only at an R&D cost of c_i. The sequentiality is translated by the assumption that innovation 2 can only be achieved by using the results of innovation 1; therefore, if the first innovation does not exist, nor does the second. We assume that it is socially desirable to produce the two innovations: $v_1 + v_2 > c_1 + c_2$. The question is how to design patents so as to reach the social optimum. One could think first of granting a patent for each innovation. For this option to work, it must be that $v_1 > c_1$. Otherwise, firm 1 will not find it profitable to invest in R&D and neither the first nor the second innovation will be produced.

In the latter case (i.e., if $v_1 < c_1$), we need to consider the alternative option of granting a broader (or 'deeper') patent to firm 1 so as to provide it with the incentive to produce the first innovation. This means that firm 1 would have rights on innovation 2 as well, but such a situation would put firm 2 at risk. Indeed, firm 2 can rationally anticipate being held up: once it has sunk the R&D cost of producing the second innovation, it will be in firm 1's best interest to appropriate the total value of this innovation, v_2. Therefore, firm 2 will not be able to recoup its investment and will prefer not to invest. As a consequence, innovation 2 will not be produced. Moreover, as $v_1 < c_1$ and as v_2 does not materialize, firm 1 will have no incentive either to invest in the first innovation. We can therefore conclude the following.

Lesson 19.7 **Because of the holdup problem, a broad patent covering later developments of an innovation is not more efficient than a sequence of narrow patents.**

In a more general setting, however, it can be argued that, because of the difficulties in dividing profit, patent lives will have to be longer than if the whole sequence of innovations occurs

[36] We follow here the simplified version of the Green and Scotchmer (1995) model presented by Lévêque and Ménière (2004).

in a single firm.[37] Then *ex ante licensing* (i.e., licensing before the second innovator sinks funds into R&D) is a way of mimicking the latter outcome. The other possibility is that licensing occurs *ex post*, that is, after the second innovator has achieved the improvement or application of the first innovation. The difference, of course, is that the second innovator has sunk his costs at the time of the negotiation, which opens the door for opportunistic behaviour on the part of the first innovator (i.e., of holdup).[t]

In contrast, it can be argued that in the case of sequential innovations, patent protection is not as useful for encouraging innovation and might even be counterproductive.[38] This can be shown in a sequential model of innovation in which an innovator's prospective profit may actually be enhanced by competition and imitation. The idea is that the loss in revenue due to increased competition may be offset by the long-term gain of being able to share the available technologies. Open-source software is an example that seems to fit the predictions of this model. We return to this topic in Section 19.4.

Patent assertion entities

The term 'patent assertion entities' (PAEs) is used to describe companies that acquire patent portfolios not to exploit them but to negotiate licences, potentially under the threat of an action for patent infringement. Such companies are sometimes identified as 'patent licensing and enforcement companies', or as 'non-practicing entities'. They also go by the more disparaging name of 'patent trolls'. Trolls are small evil creatures from Norse mythology, which often live under bridges and pester travellers for safe passage; similarly, PAEs are accused of taking firms by surprise once they have made irreversible investments.

Component-driven industries, like information technology, are particularly prone to such holdups. The problem is more acute in the USA, where patent infringements can be punished through a permanent injunction (i.e., a court order against further infringement). PAEs can then use the threat of permanent injunction to extort hefty fees in licensing negotiations, or huge settlements from companies they have accused of infringing. A famous example is the $612.5 million out-of-court settlement that Research in Motion accepted paying to NTP in 2006 to avoid the risk of its popular BlackBerry service being shut down. Because such tactics may stifle innovation in the long term, the US Supreme Court decided in 2006 to review the practice of automatically issuing a permanent injunction whenever a patent was found valid and infringed. Furthermore, in 2012, the America Invents Act (which significantly reformed the US patent system) provided additional safeguards against the potential extortionist behaviour of PAEs. For instance, the new law limits PAEs' ability to join multiple defendants in a single lawsuit, thereby increasing their costs.

Up to 2013, the European market was not PAEs' favourite playground. However, a number of large companies (including Adidas, Deutsche Telekom, Apple, Google and Microsoft) expressed

[37] See Green and Scotchmer (1995).
[t] We expect thus more licensing agreements occurring ex ante rather than ex post in industries which are known for sequential innovation. Yet, empirical evidence suggests otherwise: Anand and Khanna (2000) found that only 5% to 6% of licensing agreements occurred ex ante in computer and electronic industries, and 23% in chemicals and pharmaceuticals. Bessen (2004) suggests that the reason could be asymmetric information: if there is sequential innovation, it is probably because the second innovator possesses private information about the profitability of a subsequent innovation that the first innovator did not have (or did not find it profitable to acquire).
[38] See Bessen and Maskin (2009).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

fear that the establishment of the system of a unitary patent and unified patent court throughout the European Union could make patent assertion more profitable in Europe. Their argument has to do with the increased leverage of the future patent: it will indeed be possible to obtain a legal injunction valid on a large market (at least 13 Member States, including Germany, France and the UK). It is not clear whether such fears are legitimate. First, the 'sword of Damocles' of an injunction already exists in the current European system. Second, in Europe, only professional judges rule on patent cases, which reduces uncertainty compared with the USA, where a majority of cases are judged by a jury. Finally, the damages that are awarded at the end of the proceedings for infringement are much lower in Europe.

Many observers disapprove of the actions of PAEs because they assimilate their business model to judicial 'blackmail'. Yet, PAEs also have their proponents, who consider that these companies do nothing wrong: they simply enforce a legal right that they own. PAEs may also contribute to make the market for patents more liquid and more efficient. There are indeed two main reasons why markets for patents work badly. First, many uncertainties surround the value of patents; this is mainly because patents are idiosyncratic (and thus escape any type of metrics) and because they often need to be combined with other patents to create any value (a so-called 'portfolio effect', which we describe in the next section). Second, search costs are very high; that is, potential buyers and sellers have a hard time finding the right trading partner on the other side of the market. PAEs, as their proponents argue, may help solve these market failures.[39]

Large innovating companies also seem to have an ambivalent attitude to PAEs: officially, they complain about PAEs' alleged extortionist behaviour; yet, they sometimes hire PAEs for undercover operations of patent assertation, as explained in Case 19.5.

Case 19.5 Patent privateers

In 1572, Queen Elizabeth I of England commissioned Francis Drake to sail for America and encouraged him to plunder Spanish vessels on the way. Drake had thus all the attributes of a pirate, except that he was not working on his own account but for a government. This type of pirate was known as a 'privateer'. Privateers presented the big advantage of allowing one nation to harry another without officially attacking it. In other words, nations were avoiding the costs (and risks) of state-run warfare by outsourcing the 'business' to profit-maximizing entrepreneurs.

Although privateers disappeared with naval warfare in the 19th century (and an agreement by all major European powers to abolish it), they seem recently to have reincarnated into a new breed called 'patent privateers'. Like their forefathers, patent privateers are armed by a powerful sponsor, with the aim of assaulting their sponsor's rivals. In the 21st-century version, sponsors – and their rivals – are established operating companies, mainly in the IT and consumer electronic sectors. The arsenal also differs from the 16th-century version: instead of ships and guns, privateers receive patents and raids take the form of expensive and incessant patent infringement litigation.

[39] For a description of intermediation in patent markets, see Hagiu and Yoffie (2013). For an empirical study of how the market for patents affects enforcement of patent rights, see Galasso, Schankerman and Serrano (2013).

Here are two examples of patent privateering agreements: the first took place in 2011 between Microsoft, Nokia and MOSAID; the second was concluded in 2013 by Ericsson and Unwired Planet. Both MOSAID and Unwired Planet (formerly known as Openwave Systems) are well-known PAEs, and both deals involved the transfer of more than 2000 patents.

Sponsors of privateers see nothing wrong in such deals: they are simply looking for the most efficient way to realize a legitimate return on their high and risky investments in intellectual property. Using privateers is efficient indeed as it allows operating companies to outsource litigation and to avoid a countersuit against their own operations; and if the competitors' costs are raised in the process, even better! Naturally, the targets of the privateers' raids have a totally different opinion. Standing at the forefront of the fight against privateering is Google. In April 2013, Google wrote a letter (co-signed by BlackBerry, Earthlink and Red Hat) to the Federal Trade Commission and the US Department of Justice to ask them to take action not only against PAEs but also against the companies who supply them with weapons.

19.3.2 Complementary innovations and anticommons

If innovations are complementary, the patent system creates what Shapiro (2001) calls a *patent thicket*: 'an overlapping set of patent rights requiring that those seeking to commercialize new technology obtain licenses from multiple patentees'. The fear is that, because it is costly for firms to 'hack their way' through this dense thicket, stronger patent rights can have the perverse effect of stifling, not encouraging, innovation. We first explain the origin of this problem and discuss how serious it can be. We then turn to potential solutions. Finally, we discuss how firms' reactions to this problem may end up amplifying it.

Pricing of complements

The patent system could be in danger of imposing an unnecessary drag on innovation by enabling multiple rights owners to 'tax' new products or processes. This problem is well understood since Cournot's classic work on the pricing of complements in 1838. Cournot showed that a single monopolist of several complements sets a lower price than separate monopolists, each controlling one of the goods. The intuition behind this result is simple: each individual firm ignores the positive effect that a decrease in its own price has on the demand for the other firms' products; in contrast, an integrated firm internalizes the complementarity between the products and, hence, has a further incentive to decrease prices. This situation is sometimes referred to as the *tragedy of the anticommons*, to describe the fact that when several individuals own rights of exclusion and exercise those rights, they restrict access and therefore use of common resources.[40] This is exactly what applies when multiple firms control essential patents for a new product or process: separate licensing results in higher total prices due to multiple marginalization (or 'royalty stacking'). As the transaction costs of contracting with each patent holder also pile up, the total costs of the

[40] Heller and Eisenberg (1998) explain why the classic complements problem of Cournot can be seen as the mirror image of the well-known tragedy of the commons (which refers to the fact that a resource – like fishing grounds or clean water – can be overused if it is not protected by property rights).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

new product or process necessarily increase. In the next case, we describe an attempt to estimate empirically the seriousness of this problem.

Case 19.6 Do patent rights facilitate or impede cumulative innovation?

Galasso and Schankerman (2013) estimate the impact of patent rights on cumulative innovation. Such estimation raises three empirical difficulties. First, one needs to identify comparable technologies with and without patent protection. The author address this issue by exploiting patent invalidation decisions by the US Court of Appeal for the Federal Circuit (which has exclusive jurisdiction in appellate cases involving patents); their sample comprises 1357 decisions from 1983 to 2008, with about 40% of these decisions having led to a loss of patent protection for the technology. The second empirical difficulty is that subsequent innovations are difficult to identify. Here, the authors follow the strategy that is commonly used in the literature: they exploit citations by later patents. As we discussed in Section 19.1, a patent can only be granted if it is deemed sufficiently novel with respect to the prior art; patent applicants have thus to describe the prior art and for that matter, they have to cite all relevant patents, even those that have been invalidated.

The last and major challenge is endogeneity: the factors that make invalidated patents differ from those that are upheld may also affect patent citations; for instance, the commercial potential of a patent may be a common explanation for the facts that this patent led to subsequent innovations and that its owner fought hard to avoid invalidation. To circumvent this problem, the authors exploit a couple of institutional facts: judges are randomly assigned to patent cases, and they form panels of three that decide by majority voting. Because judges differ in their propensity to invalidate patents, the author can use these facts to construct an instrumental variable that addresses the potential endogeneity of invalidation decisions.

It is thus possible to estimate the impacts of patent rights on cumulative innovation by examining how invalidation of a patent affects the rate of subsequent citations to that patent. The main result is that the removal of patent protection leads to about a 50% increase in subsequent citations to the focal patent. This lends credence to the idea that patent rights do impede cumulative innovation. However, the authors stress the facts that the 50% increase is an average estimate, and that large variations are observed across sectors. Actually, the positive impact of patent invalidation on subsequent innovation is significant (and large) only in sectors with complex technology (namely computers and communications, electronics and medical instruments). This is not a surprise as these sectors are also those where the patent thicket is particularly dense. The authors also observe variations across firms along the size dimension: it is mainly the invalidation of patents held by large firms that triggers subsequent innovation by small firms.

Patent pools and cross-licensing

What Cournot (1838) shows when analysing the pricing of complements is that not only the consumers but also the firms benefit from the ‘collusive’ pricing of complements. Therefore, right holders will find it profitable to coordinate their decisions, either by creating a *patent pool* or by engaging in *cross-licensing*. Under a patent pool, an entire group of patents is licensed in a

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

package, either by one of the patent holders or by a new entity established for this purpose, usually to anyone willing to pay the associated royalties. A cross-licence is simply an agreement between two companies that grants each the right to practise the other's patents (possibly involving fixed fees or running royalties).[u]

These forms of coordination, however, entail transaction costs and might be seen with suspicion by antitrust authorities because of price effects. Regarding antitrust concerns, the previous argument suggests that package licensing is desirable for complementary patents but not for substitute patents. It can be shown that requiring pool members to be able to independently license patents matters if and only if the pool is otherwise welfare-reducing, a property that allows the antitrust authorities to use this requirement to screen out unattractive pools.[41]

Even when patent pools are welfare-enhancing, their formation is far from granted. First, we recall from Chapter 14 that cartels are inherently unstable: although they are collectively profitable, each member has an individual incentive to cheat on the agreement. The same goes for patent pools: each patentee has an incentive to free-ride on the other pool members' efforts to lower the total price of the complementary patents, and to increase the price of its own patent. A second source of instability is the potentially large asymmetry across the patents constituting the pool, and across the patent holders themselves. As far as patents in the pool are concerned, technological complexity contributes to increase not only the number of complementary patents but also their heterogeneity: they may have involved very different R&D costs for their owners, and they may cover very different technologies (breakthroughs or mere improvements). Regarding patent holders, patent pools may unite organizations with conflicting objectives (e.g., private firms seeking revenue maximization, and public institutions concerned by the diffusion of a standard) or with conflicting interests (e.g., pure patent holders, which derive revenue solely from licensing, and integrated firms, which also have to pay royalties for their manufacturing activities).[42] Finally, the formation of patent pools may be jeopardized by the prospect of a form of holdup called 'patent ambush'. The risk is particularly accute for patent pools negotiated within standard setting organizations (SSOs): a patent holder may first mislead the SSO into believing that it does not hold patents in essential technologies; it would then wait until the SSO includes the unknowingly patented technology in industry standards, and finally reveal patent ownership by suing the pool members for infringement. Such ambush is particularly attractive when the standard in question is likely to become dominant; then, its users will be locked in and hence, will not be in a position to negotiate a reasonable royalty rate. To reduce the risk of patent ambush, most SSOs enjoin their members first to identify and disclose any patent that may be relevant for the standard under development, and second to license the disclosed patents on Fair, Reasonable And Non-Discriminatory (FRAND) terms.[v] The FRAND requirement may also reinforce the stability of the pool by ensuring a fair sharing rule of the pool revenues. However, the main weakness of the

[u] An example of a patent pool is the MPEG-2 video compression technology. Nine companies have pooled their patents to permit one-stop shopping for makers of televisions, digital video disks and players and telecommunications equipment as well as cable, satellite and broadcast television services. Shapiro (2001) reports that broad cross-licences are the norm in markets for the design and manufacture of microprocessors.

[41] This issue is formally studied by Lerner and Tirole (2004), who provide a necessary and sufficient condition for a patent pool to enhance welfare. Rey and Tirole (2013) address the same issue in the context of repeated interaction and coordination. Here, a patent pool affects the firms' ability to tacitly coordinate their pricing decisions.

[42] In the context of voluntary standard-setting organizations, Simcoe (2012) finds evidence that standards are delayed because of distributional conflicts among patent holders.

[v] The acronym FRAND is sometimes written (F)RAND to account for the fact that the fairness requirement is more commonly applied in Europe than in the USA.

concept lies in its vagueness, especially regarding the FR part, whose interpretation is often left to courts and regulators. Whether the FRAND requirement provides an effective protection against holdup remains thus an open question.[43]

Patent portfolios

It appears from the previous discussion that in industries relying on cumulative innovations, firms seek to constitute portfolios of related patents, because the aggregation of these patents gives to the portfolio a value that is larger than the sum of the values of the individual patents. In other words, the more patents the merrier, because patent portfolios can be used as a 'trading device' or a 'bargaining chip'. Indeed, in the presence of patent thickets, a firm is very likely to infringe another firm's patent; then, a large portfolio provides the infringing firm with a credible threat of counter-infringement litigation. A large portfolio also improves its holder's bargaining position with other firms.

We can find here an explanation for the patent paradox that we mentioned above (i.e., the combined facts that patent counts have exploded over the last decades, while innovators view patents as a secondary way to protect their investments in R&D).

19.4 Intellectual property in the digital economy

The main form of IP protection that we have considered so far in this chapter is the patent system. The patent system is designed to protect industrial property, such as inventions, processes, machines, etc. Yet, as we mentioned previously, there is a wide range of other productions of information which suffer from the same appropriability problem as industrial innovations, but which are not covered by the patent regime. We think of the production of literary, musical, choreographic, dramatic and artistic works (such as novels, poems, plays, films, songs, drawings, paintings, photographs, sculptures, architectural designs, etc.). Such works are protected by *copyright*. Although copyrights and patents stem from the same economic rationale (absent legal protection, creators lack incentives to engage in the production of information), they strike a different balance between static and dynamic efficiency considerations: roughly put, the protection offered by copyright is longer but narrower than the protection offered by a patent. Copyright applies to the expression of works, in whatever mode or form, and gives authors an exclusive right over the reproduction, performance, adaptation and translation of their work. Compared with patents, this protection is weaker (as only the expression is protected, and not the underlying ideas) but it is extended over a longer period of time (nowadays, it lasts for 70 years after the author's death in both Europe and the USA).

Most of the arguments we have developed above remain useful in understanding the economics of copyright. However, the industries producing copyrighted goods, and the way people use and enjoy these goods, have recently been deeply altered by digital technology and the

[43] Some recent papers shed light on this issue. Dewatripont and Legros (2013) analyse how a FRAND-like setting affects incentives to invest in R&D and the quality of the patents that firms contribute in a standard. Tirole and Lerner (2013) build a framework for the analysis of standard-essential patents (i.e., patents that are ex ante not that important but become essential once they are included into a standard). Llanes and Poblete (2014) study the effects of alternative standard-setting and pool-formation rules on technology choice, prices and welfare.

Internet. Because it modifies the interaction between copyright holders, technology companies and consumers, this evolution poses interesting challenges for the economic analysis of IP protection.

In this section, we build on our previous analysis and develop some new tools in order to address topical issues. We focus on music and software as these two industries have experienced the most significant change and so have dominated the public debate. Regarding the music industry, we examine the impact of end-user copying on the decisions of copyright holders, as well as on total welfare. Regarding the software industry, we discuss the appropriate IP regime for software, and we examine the development of open-source software.

19.4.1 End-user piracy

IP laws are effective only if they are properly enforced and respected.[44] Yet, as far as information products are concerned, one observes a large-scale violation of the laws protecting them, a phenomenon known as 'piracy'. What may appear striking is that the illegal reproduction and distribution of copyrighted works is not only the act of criminal organizations (so-called *commercial* piracy) but also the act of consumers themselves (so-called *end-user* piracy).

Commercial piracy does not need much analysis, as the motivation is easily understood: criminal organizations are simply attracted by the high profit margins that the large-scale reproduction and distribution of copyrighted products generates. On the contrary, end-user piracy raises a number of issues that the fast penetration of the Internet and the digitization of information products have made much more pressing. Digital technologies have indeed drastically reduced the cost of making and distributing illegal copies, while increasing their quality; thereby, they have deeply modified the interaction between end-users, copyright holders and technology companies. End-user piracy in the digital age, or for short digital piracy, is thus a major phenomenon that requires some analysis.

The basic economic analysis: digital piracy decreases profits

The main consequence of digital piracy is that it seriously limits copyright owners in their ability to control how information products get to consumers. As a result, the availability of digital copies is likely to reduce the copyright owner's profits. This is the prediction that can be drawn from the basic theoretical modelling of piracy.[45] These models simplify the analysis by focusing on the market for a digital product supplied by a single producer.[w] Even though the copyright owner acts as a monopoly, she faces nevertheless the competition exerted by the availability of (illegal) digital copies. Copies are seen as imperfect substitutes for the original digital product, insofar as their quality is generally lower than the quality of original products. In particular, the quality of copies primarily depends on technological and legal factors, which can be affected by public authorities (through the definition and enforcement of IP protection) and/or by the copyright owner herself (through technical protective measures). In this setting, it is possible to analyse the copyright owner's decisions about the pricing and the technical protection of original products, as well as public policy regarding IP laws.

[44] This section takes some extracts from Belleflamme and Peitz (2014).
[45] See, for instance, Novos and Waldman (1984), Johnson (1985) and the references in Belleflamme and Peitz (2012).
[w] One can justify this assumption by arguing that digital products within a given category are highly differentiated in the eyes of the consumer; the demand for any product is therefore hardly affected by the prices of other products.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

The main results of these analyses can be summarized as follows. First, because consumers with a low cost of copying or with a low willingness to pay for quality prefer copies to original products, the copyright owner is forced to charge a lower price (than in a world where digital piracy would not exist). That clearly decreases the copyright owner's profits but increases the surplus of the consumers of original products; moreover, a number of consumers who were not willing to purchase the original product at the monopoly price now get some utility from the pirated copies. As the increase in consumer surplus outweighs the profit reduction, digital piracy results in an improvement of welfare from a static efficiency point of view (like any erosion of market power does). However, the lower profits may reduce the incentives of copyright owners to improve the quality of existing products or to introduce new products on the market; this is detrimental to welfare from a dynamic perspective. Moreover, total welfare may decrease because of a number of avoidable costs that digital piracy entails (e.g., the costs for producers to implement technical protective measures, or the costs for public authorities to enforce copyrights).

Looking at the profits of copyright owners, it is an undisputed fact that they started to decrease when end-user piracy started to grow (i.e., around 1999 with the launch of Napster, a peer-to-peer file-sharing service). This was particularly acute in the music industry, where physical music sales (that is to say, CDs) dropped significantly. Numerous empirical studies have tried to estimate the extent to which this decrease in sales could be attributed to digital piracy. These studies converged on the conclusion, now widely accepted, that digital piracy has 'displaced' physical sales (i.e., legal purchases were substituted for, mainly, illegal downloads). However, it is also established that the estimated 'displacement rate' is slightly above zero and nowhere near unity, reflecting the observation that the vast majority of goods that were illegally consumed would not have been purchased in the absence of piracy (contrary to what the recording industry would have liked the general public to believe by counting any download as a lost sale).[46]

Further developments: digital piracy may increase profits

A number of theoretical studies have demonstrated the positive effects that piracy may have on the profits of copyright owners. Three mechanisms have been identified.[47] First, illegal copies of a digital product can play a sampling role by attracting consumers and driving them to purchase a legitimate copy later. This argument is based on the observation that digital products are complex 'experience goods'; that is, consumers do not know the exact value that they attach to particular digital products before consuming them. Buying a legitimate copy may thus appear risky, which inevitably reduces demand. However, if an illegal copy can be accessed free of charge, consumers may learn their valuation of the product and if the latter is large, they may want to purchase the legitimate product (which is often, as argued above, of a higher perceived quality).

[46] For a survey of the empirics of digital piracy, see Waldfogel (2012a). Very little empirical work has been devoted to the long-term effects of piracy (i.e., to dynamic efficiency considerations). One notable exception is Waldfogel (2012b), who tries to estimate the extent to which digital piracy has affected the incentives to bring forth a steady stream of valuable new products. To address this issue, he uses three different methods to assess the quality of new recorded music since Napster. The three resulting indices of music quality show no evidence of a reduction in the quality of music released since 1999; two indices even suggest an increase. One explanation could be that the digital technologies that have made piracy easier have also reduced the costs of bringing creative works to market, and the latter effect is at least as important as the former.

[47] See Peitz and Waelbroeck (2006) and the references in Belleflamme and Peitz (2012).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

The second mechanism originates in the fact that many digital products generate network effects (see Chapter 20 for the definition and analysis of network effects); that is, the attraction of the product increases with the number of consumers of that product. This is so with software (the wider the community of users, the easier it is to exchange files and the larger is the supply of complementary products) or with cultural products (whose popularity increases with word-of-mouth). As it is the cumulated number of consumed copies that matter and not whether these copies are legitimate or not, digital piracy contributes to increase the willingness to pay for legitimate copies.

Finally, the third mechanism, called indirect appropriation, resembles the second by invoking the fact that piracy can increase the demand for goods that are complementary to the pirated content; the producer is then able to capture indirectly the value that consumers attach to the pirated good. This goes, for example, for increasing ticket sales for the concert of an artist whose popularity may be due partly to a large base of fans consuming pirated copies of this artist's songs.

The following case illustrates how these three mechanisms may explain why digital piracy does not seem to worry too much the producers of a popular TV series.

Case 19.7 The hidden treasure of piracy?

You probably know the TV series *Game of Thrones*. The first two seasons were a huge hit and the third season premiered on 31 March 2013 on HBO (the American premium cable network that produces the series). There are about 4.5 million subscribers who follow the series. But overall, the audience is much broader as one should also count several million (the exact figure is difficult to estimate) of 'illegitimate' viewers. For example, the first episode of Season 3 was downloaded over one million times in the space of 24 hours.

Contrary to what one might think, these record numbers of illegal downloadings have not panicked HBO officials. Quite the opposite! HBO programming president (Michael Lombardo) told *Entertainment Weekly* (on 31 March 2013) that he viewed this rather positively: 'I probably should not say this, but this is a kind of compliment. The demand is there. And it clearly did not have a negative effect on DVD sales.' One of the directors of the series (David Petrarca) went one step further by stating that, in fact, the series benefits from piracy because it feeds the 'cultural buzz' that allows this kind of programme to 'survive'.

This tolerant attitude can be explained by the three mechanisms described above. First, the sampling effect seems to play a role as it is estimated that a large number of viewers watch the series first through illegal downloads and then again on cable TV or on Blu-Ray DVDs. Second, network effects follow from word-of-mouth: the more the series is viewed, be it through legal or illegal channels, the more it is talked about and the more HBO is likely to attract new subscribers. Finally, the popularity of the series also boosts demand for the associated merchandise, for which hardcore fans have a high willingness to pay.

Perspectives

The presence of these potential positive impacts of piracy and the inability to preserve the existing business models drove the content industries to experiment with new solutions. Because it had been the first to be hit by digital piracy, the music industry also took the lead in terms of innovative

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

business models. The first answer to falling CD sales was to move the distribution of music online. At the forefront was the iTunes Music Store operated by Apple, which opened in 2003. These legal online channels for digital music allowed consumers not only to find and download music as easily as via illegal channels, but also to start buying individual tracks instead of being forced to buy albums. This may explain why online legal sales of digital music seem to suffer less from piracy than physical sales do.[48]

New business models in the music industry also offer market solutions to increase revenues from the segment of consumers with a low willingness to pay for music and with, therefore, a high disposition to digital piracy. Streaming services (such as Spotify or Deezer) are based on a 'freemium' model, which combines free and premium (i.e., paying) services. The objective is to attract users with the free offering and, later, 'convert' them to paying subscribers. This objective can be reached through different ways: the premium offering can include additional 'mobility' (e.g., the possibility to access playlists on various devices, such as a computer, tablet or smartphone), better sound quality, a wider library of titles or the removal of ads.

Markets for information products are undergoing major changes due to technological innovations, which triggered digital piracy and, partly as a response, new business models. As exemplified above, in this changing landscape, some research suggests that consumer behaviour exhibits several interesting features. Whether these features are stable over time and space is an interesting area for future research. Such an understanding is necessary to evaluate the impact of digital piracy on markets for information products and to develop successful new business models. It is also necessary to propose appropriate public policy responses. This is all the more important as digital piracy may soon extend to physical objects, as the next case suggests.

Case 19.8 3D printing piracy[49]

3D printing is the process of making three-dimensional solid objects from digital designs. From scale models, gifts and clothing to prosthetic limbs, the practical applications are multiple and seem limitless. The industrial use of 3D printing is not new: several industries have been using additive manufacturing (i.e., the creation of objects through a sequential layering process) for quite a long time. The main applications are rapid prototyping, rapid manufacturing and customization. Compared with traditional manufacturing technologies, additive manufacturing implies a different cost structure (arguably, lower economies of scale and larger economies of scope) and an increased ability to meet demand. What is new though (since the early 2010s) is the widespread use of 3D printing among individuals. This is made possible by the constant improvement of the technology, associated with a significant fall in prices. It thus seems legitimate to study what would be the consequences of letting consumers produce a great number of objects by their own forces. Will 3D printers challenge manufacturing industries in the same way as digital piracy has challenged the music industry?

[48] Koh, Murthi and Raghunathan (2013) show that it is the legal sales of online music and not digital piracy that displaced physical music sales after 2003. In the same vein, Aguiar and Martens (2013) conclude that the online legal sales of digital music (through online stores such as iTunes or via streaming services such as Spotify) do not seem to be displaced by illegal downloading; the opposite may even occur.

[49] This case is based on Simonite (2013).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

The unauthorized copying and sharing of designs is indeed a technical possibility. Even though this possibility has not materialized yet, commercial solutions are already proposed to address the problem. For instance, the startup Authentise has designed a DRM-type software that makes it possible to send a design to a 3D printer and limit its printing to a single copy.[p]

19.4.2 Software protection

Although the subject matters of patents and copyright are quite distinct, some innovations have elements that pertain to the two categories and it is therefore not clear whether such innovations should benefit from the protection of patents or of copyrights, which differ in terms of length and breadth. The typical example is software. Software has traditionally been protected by copyright but since the early 1970s, the US Patent and Trademark Office (USPTO) broadly grants patents that may be referred to as software patents. In contrast, the European Patent Office (EPO) states that 'a computer program claimed "as such" is not a patentable invention. Patents are not granted merely for program listings. Program listings as such are protected by copyright. For a patent to be granted for a computer-implemented invention, a technical problem has to be solved in a novel and non-obvious manner.'

The extent to which software should be patented has been – and still is – a topic of intense debate. To shed some light on this debate, it is useful to examine the economic specificities of software, and to assess whether the 'one-size-fits-all' tool of patent protection is appropriate given these specificities.

Economic specificities of software

Producers of software have to be protected not only from competitors' imitation, but also from consumers' copying. The protection against consumers' copying is the economic role of copyright. As for the protection against imitation by competitors, patents seem more appropriate. Yet, considering that software innovation involves relatively low costs compared with the private and social returns it generates, patents might confer too strong a protection from a social point of view.

The previous conclusion has to be qualified when we take into account the intrinsic *cumulativeness* of software innovation (as discussed in Section 19.3). Software innovation exhibits both types of cumulativeness: sequentiality and complementarity. First, given that any new software is built upon previous lines of code, innovation is sequential, which calls for broad (and short) patents. Second, because there are complex interdependencies between different software and interoperability is very often a major concern, software innovations are complementary. This raises the prospect of the 'tragedy of the anticommons', when strong property rights are awarded to separate right holders. In this respect, protection should be designed optimally so as to favour interoperability, with easy access to interfaces.

Another important feature of software is the presence of *network effects*. There are various reasons for which software becomes more valuable for its users the more widely it is adopted.

[p] Digital rights management (DRM) systems inhibit uses of digital content not desired or intended by the content provider. DRM systems are meant to fight digital piracy but also, more generally, to manage how digital products can be used.

As we will explain in Chapter 20, the presence of such reinforcement mechanisms drives the competition between incompatible software towards a 'winner-takes-all' scenario: it is very likely that a single software will end up dominating its market. In such a context, one might question the necessity of supplementing this natural tendency towards monopolization with the creation of a legal monopoly resulting from strong IP rights.

An opposite argument can be proposed when invoking the *durability* of software. As software does not wear out, we recall from Chapter 10 that, according to the Coase conjecture, a software producer might not be able to price above marginal cost because it cannot commit not to decrease prices in the future. Therefore, a stronger IP protection should be granted to software producers so as to provide them with the incentives to create. However, software producers have designed tactics to counter the durability problem. For instance, producers can sell subscription instead of standalone software packages. Software is then distributed 'as a service' rather than 'as a product': instead of letting users purchase the software and install it on their personal computer, the software is hosted by the vendor and made available to users over the Internet. Software producers can also add new features that improve upon older versions in order to create a new flow of demand. Another commonly used tactic consists of 'planning the obsolescence' of older versions by decreasing compatibility and/or technical support.

Optimal protection?

Collecting the previous arguments, we realize how complex it is to assess whether patent protection should be extended to software. On the one hand, we understand how software patents could foster innovation by increasing private returns on R&D. On the other hand, we also realize that software patents could stifle innovation because they grant too strong a protection given the sequential and complementary nature of software, and because of the heavy need for interoperability. In one of the rare empirical analyses on this topic, Bessen and Hunt (2007) argue that regulatory changes that reduced the cost of software patents in the USA led firms to patent software innovations more for strategic motives than for covering R&D expenditures. Their analysis also indicates a negative correlation between software patenting intensity and R&D intensity in the software sector.

The previous discussion and the contrasting approaches of the USPTO and the EPO are in terms of 'patentability' (or patent-eligible subject matter): what are the conditions for a software (or a 'computer-implemented invention' in the European terminology) to be patentable? Two other, and potentially complementary, ways to tackle the issue would be to discuss the 'patent length' and the 'patent breadth' for software. In terms of length, some scholars recommend a shorter patent term for digital technologies. Other scholars suggest reconsidering the breadth of software patents. For instance, Lemley (2012) suggests limiting software patents to their specific way of accomplishing a function; more precisely, he recommends preventing the practice of 'functional claiming', which consists of patenting a software function rather than a specific way to implement that function. In other words, protection should be awarded only to an innovator's solution to a given problem and not to the problem itself.

No protection: open-source software

Software can be transmitted in either *source code* (i.e., code using languages such as Basic, C and Java, which can easily be interpreted and modified by programmers) or *object code* (i.e., sequence of 0s and 1s that communicate directly with the hardware). Most commercial software vendors provide users only with object code.[q] In contrast, open-source software (OSS) is, as its name indicates, such that the source code is made available to everyone. While open-source presents a way to broadly make the sources of a product publicly accessible, open-source licenses allow the authors to fine tune such access by defining users' rights on usage, modification and redistribution.[r]

Major OSS projects are nowadays competing head-to-head with commercial software vendors (e.g., Linux, Apache or Firefox, on the markets for, respectively, server operating systems, web servers and web browsers). The successful development of OSS projects puzzled economists. At first glance, economists would not expect that complex software could be produced through a loose form of coordination between a very large number of individual programmers, who donate their time and effort free of charge. However, this is exactly how Linux, Apache and many other OSS projects have been, and still are, developed. Consequently, a substantial and growing economics and business literature has addressed the organizational, motivational and product-structural features of OSS development.[50]

We briefly examine here the roles and incentives of the various actors in the open-source process.

Motivations of open-source contributors. Why do top-notch programmers choose to write code that is released for free? Is such behaviour consistent with the self-interested economic-agent paradigm? A simple cost/benefit analysis allows us to understand better what motivates programmers to participate in OSS projects. On the *cost side*, the programmers incur an opportunity cost of time (which can be measured by the monetary compensation that they could earn were they working instead for a commercial software vendor). On the *benefit side*, several short- or long-run benefits may outweigh these costs. First, many contributors are sophisticated users who need to remove a bug or tailor the code to their specific applications; contributing to an OSS project may thus improve their performance in paid work. Second, having turned the code back to the community, they may see others improve on their modifications, increasing their private benefit further. Third, programmers may find more intrinsic pleasure and fun in contributing to a 'cool' OSS project rather than working on a routine task set by an employer. Fourth, the delayed benefits are likely to be non-negligible. Contributions to OSS projects are indeed well recognized among peers; this provides contributors not only with ego gratification, but also with the prospect of future monetary compensation (in the form of better job offers, shares in commercial OSS companies or future access to the venture capital market).

Finally, comparing the commercial and open-source environments, it is fair to say that the former offers better current compensation, while the latter generates larger delayed rewards. There are indeed three reasons why programmers find it easier to signal a high level of competence

[q] IP law does not require disclosure of source code: copyrights for software can be registered without fully revealing the source code and software patents typically do not include the source code.

[r] There exist a large variety of open-source software licences. The most prominent example is the popular GNU General Public Licence (GPL), which requires that all enhancements to the original material must also be made publicly available. Note that OSS must be distinguished from freeware and shareware (which can both be downloaded free of charge but do not allow access to the source code) and from public domain software (which is not licensed and is thus usable by everyone without constraint).

[50] See Johnson (2012) for a comprehensive literature survey. We follow mainly Lerner and Tirole (2001, 2002, 2005a,b) and Maurer and Scotchmer (2006).

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

in the open-source environment: (i) the contribution of each individual can be seen by a large community of outsiders; (ii) the programmer takes full responsibility for the success of a subproject; (iii) because of the cumulative nature of software innovation, knowledge accumulated on a particular subproject can easily be transferred to another project, which increases the programmer's attractiveness for future employers.

Attitude of commercial firms. The way for-profit firms work and compete with OSS projects is much less intriguing for economists than the 'volunteer' participation of programmers. First, a profitable business model consists of developing a proprietary market segment by offering specific expertise or support services that are complementary to open-source software (which is not appropriable in itself). For instance, Red Hat packages Linux in a user-friendly way and provides additional material such as manuals and support services. Second, commercial companies tend to allocate some of their talented staff to open-source programs. This allows them to keep abreast of open-source developments and hence know the competition and develop an absorptive capacity (for incorporating open-source ideas into commercial software and for spotting talented programmers). Third, commercial firms may compete directly with OSS providers in the same market. In contrast, they may participate in an OSS project for strategic reasons: embracing an OSS project may pre-empt the development of a standard around a technology owned by a powerful rival.

Furthermore, commercial software vendors may find it profitable to release some existing proprietary code, and then rely on the open-source community for the ensuing development process. For this *code-release* strategy to make sense, the company must be able to increase profit in a proprietary complementary segment and thereby offset any profit that would have been made in the primary segment. This is more likely to happen when the company already lags behind its competitors in the primary segment, or when network effects and switching costs are strong (which implies, as we will see in Chapter 20, that the best-selling software captures almost the entire market). As software innovation is largely cumulative, code release may also induce further development, which could eventually benefit the initial innovator.

Open-source vs. traditional IP incentives. The main advantage of open-source with respect to the legal protection of IP is that it avoids the deadweight loss resulting from monopoly pricing. Open-source therefore improves static efficiency (it also accelerates the discovery of subsequent innovations through automatic disclosure). Of course, the important question is whether it does so at the expense of dynamic efficiency. As innovators do not get any (immediate) monetary reward, incentives have to be found elsewhere. We have reviewed above the entire suite of incentives that open-source comprises. What is important to note is that these various incentives have separate and distinct welfare implications: 'own use' incentives may lead to underprovision of code as benefits conferred on third parties cannot be appropriated; 'signalling' incentives may lead programmers to invest more in projects in which they can showcase their competence, rather than in projects with most consumer value; among the 'social psychological' incentives, those based on extrinsic motivations (desire for reputation within the open-source community, 'ego boost', 'feelings of personal efficacy', etc.) and on intrinsic motivations (creative pleasure, desire to be part of a team, satisfaction and accomplishment, ideological opposition to proprietary software, etc.) may lead developers to partly internalize the social benefits they confer.

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

Review questions

1. Explain how IP law strikes a balance between dynamic and static efficiency considerations or, in other words, between incentives and use.
2. What is behind the so-called ‘patent paradox’? How can it be explained?
3. Why isn’t it optimal in terms of public policy to have patents that last forever? Discuss.
4. Does a firm have incentives to license its innovation to rival firms? Discuss.
5. What is the meaning of the ‘tragedy of the anticommons’? How does this problem apply to innovations and how can it be mitigated?
6. What are the effects of end-user piracy of digital products on the producers’ choices and on social welfare? Discuss.

Further reading

The seminal paper of Arrow (1962) remains useful reading to complement this chapter. Important contributions to the optimal design of intellectual property are Nordhaus (1969) and Gilbert and Shapiro (1990). Takalo (2001) provides a nice summary of this literature. For more on the relative merits of IP protection and subsidies (and the combination of these two institutions in practice), see Chapter 8 of Scotchmer (2004). As for our treatment of IP in the digital economy, you can find more on the digital music market in Peitz and Waelbroeck (2006a) and on open-source software in Lerner and Tirole (2002, 2005b). Finally, for a less technical and more exhaustive treatment of the economics of IP, we recommend the book of Lévêque and Ménière (2004).

Exercises

19.1 Secrecy vs. patenting

Consider an innovative environment where independent or nearly simultaneous discoveries are possible. More specifically, we assume that two firms are engaged in R&D that results either in an innovation (with probability β) or failure (with probability $1-\beta$). It is assumed that the probability of success (β) is independent across firms. Firms can protect their innovation either by secrecy or by filing for a patent.

- If the innovation is protected by *secrecy*, it can leak out with probability $1-\alpha_s$. When this happens, the innovation is publicly available and production is at the competitive level, driving the innovator’s profits down to zero.

- *Patent* protection is measured by the probability that a patent holder can exclude competitors from using the innovation, which is denoted α_p. Hence, with probality $1-\alpha_p$, the innovation becomes public, resulting again in zero profits for the innovator.

If only one firm succeeds in R&D and the innovation does not become public, the firm earns monopoly profit π^m. If both firms succeed and their innovation does not become public, each firm earns duopoly profit $\pi^d < \pi^m$. In the case where both firms are successful and file for the patent, each firm obtains it with probability $1/2$.

The two firms have to decide whether to file for a patent (strategy denoted P) or resort to secrecy (strategy denoted S). This decision has to be made before learning whether the competitor has succeeded or not.

1. Using the above information, compute the firms' expected profits for the four combinations of strategies. Denote by $\Pi(a_1, a_2)$ the expected profit for a firm when it chooses strategy a_1 and its opponent chooses strategy a_2, with a_1 and $a_2 \in \{P, S\}$. You are thus asked to compute $\Pi(P, P)$, $\Pi(P, S)$, $\Pi(S, P)$ and $\Pi(S, S)$.
2. Suppose that $\alpha_p = \alpha_s$. That is, the innovation has the same probability of becoming public whether it is protected by secrecy or by a patent (in other words, patent and secrecy offer the same level of protection).
   (a) Show that patenting is a dominant strategy. That is, show that both $\Pi(P, P) \geq \Pi(S, P)$ and $\Pi(P, S) \geq \Pi(S, S)$ are true.
   (b) Show also that successful firms prefer the situation where they both file for a patent over the situation where they both keep the innovation secret. That is, show that $\Pi(P, P) \geq \Pi(S, S)$ is true.
3. Suppose now that $\alpha_p \neq \alpha_s$. To ease the computations, set $\pi^m = 16$, $\pi^d = 4$ and $\beta = 1/2$.
   (a) Compute the values of $\Pi(P, P)$, $\Pi(P, S)$, $\Pi(S, P)$ and $\Pi(S, S)$ under these assumptions.
   (b) Characterize the Nash equilibrium (in pure strategies) of the game for all $\alpha_p, \alpha_s \in [0, 1]$. Represent graphically the characterization of the equilibrium in the plane (α_p, α_s).
   (c) Show that both firms may choose to protect the innovation via a patent even though patents offer a weaker protection than secrecy. Explain the economic intuition behind this result.

19.2 Formation of patent pools

Suppose that three firms (denoted $i = 1, 2, 3$) each own a patent that is essential to the production of a given final product. For simplicity, we assume that there is a competitive industry that produces this final product, buying and assembling the necessary components from each of these three firms. We assume that the assembly firms incur no other assembly cost in addition to (i) paying royalties for the use of the three essential patents and (ii) incurring transaction costs when inquiring about the licence fees. Regarding the latter transaction costs, it is assumed that they are inversely related to the number of different licence fees that are set by the patent holders. Patent holders indeed have the possibility to form so-called 'patent pools' whereby they coordinate their decisions to set

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

a unique licence fee that allows assembly firms to access two (or three, if all firms join) patents at once. This is modelled as follows.

- The price of the final product is denoted by p. Demand for the final product is given by $q = a - p$.
- Competition in the assembly industry therefore ensures that p is equal to the marginal cost of assembly, which depends on the patent pool that patent holders may have formed. In particular, three options are possible:

$$p = \begin{cases} r_1 + r_2 + r_3 + 3\gamma & \text{if no pool is formed,} \\ r_{ij} + r_k + 2\gamma & \text{if firms } i \text{ and } j \text{ form a pool,} \\ r_p + \gamma & \text{if the three firms form a pool,} \end{cases}$$

where $0 \leq \gamma < a/3$ is the cost per transaction, r_i is the licence fee set by firm i for accessing its patent, r_{ij} (resp., r_p) is the common fee set by the pool formed by firms i and j (resp., all firms) for accessing the bundle of patents i and j (resp., all patents).

1. Suppose that the three patent holders set their licence fee independently and non-cooperatively. Derive the Nash equilibrium in fees, compute (i) the price of the final product (denote it p_s) and (ii) the profit of each patent holder at equilibrium (denote it π_s).
2. Repeat the previous analysis by assuming that firm i and j coordinate their decisions to set a common fee r_{ij} that has to be paid for acquiring the right to use patents i and j. Firm k, in contrast, still acts separately and sets its licence fee r_k. Compute again the price of the final product (denote it p_2), as well as the equilibrium profits of the firms. Denote the profit of firms i and j, π_{in} (supposing that they divide their joint profit equally) and the profit of firm k, π_{out}.
3. Suppose now that *the three firms* form a patent pool and choose a common licence fee r_p to maximize their joint profit. Assembly firms that pay r_p have access to the whole set of patents. Derive the optimal r_p. Supposing that the pool's profit is equally distributed among the three firms, compute each individual firm's profit (denote it π_p). Compute also the price of the final product (denote it p_p).
4. Comparing your answers to the previous three questions, show that, even in the absence of transaction costs ($\gamma = 0$), the more there are patents in the pool, the smaller the price of the final product and the larger the sum of the patent holders' profits. Discuss the economic intuition behind these results.
5. Consider now the following pool-formation game. Before they set the licence fees, patent holders simultaneously decide whether to join a patent pool or not (only one pool can form). Set $a = 120$ and characterize the Nash equilibrium of this game for all admissible values of the transaction cost γ (i.e., $0 \leq \gamma < 40$). Discuss the economic intuition behind your results.

19.3 Cumulative innovations

Consider the following market structure. There are two firms, denoted 1 and 2. In the first period, firm 1 exogenously makes a discovery. If it incurs costs c_1, it can turn this discovery into a new product. A mass 1 of consumers has a valuation for this product of r_1. In the second period (no discounting), firm 2 exogenously makes a discovery if (and only if) firm 1 has developed the product in period 1. Incurring costs c_2, firm 2 can turn this discovery into a competing product for which consumers have a valuation of r_2. Assume that consumers have unit demand and buy either one unit of product 1, one unit of product 2 or nothing after the second period. Throughout the exercise, assume $A1 : r_2 > c_2 + r_1$ and $A2 : r_1 > c_1 > 0$.

1. Briefly interpret $A1$ and $A2$. Which investment decisions are made in the subgame-perfect Nash equilibrium of the game without an allocation of intellectual property rights, that is, with standard competition in period 2? Explain why this result may be inefficient from a welfare standpoint.
2. For the same parameter constellation, show that a licence fee ϕ payable from firm 2 to firm 1 for every unit of the good sold can induce a welfare-optimal allocation. (Assume the following timing of the game: first the fee ϕ is set; then firm 1 makes the investment decision; finally firm 2 makes the investment decision.) In which range must ϕ lie to be effective? What happens if it is too high/too low?
3. Now let us expand the game further in the second period. (For this last part of the problem, assume for simplicity that $r_1 = 0$.) Assume that courts only enforce firm 1's licence claims against firm 2 with probability p. Firm 2 can choose two types of monetary investment: c_2^a increases the value of the product for consumers, such that $\partial r_2/\partial c_2^a > 0$, with $r_2(0) > r_1$ and $\partial^2 r_2/\partial {c_2^a}^2 < 0$. Parameter c_2^b does not affect the value of the product to consumers, but it reduces the probability that firm 2 would be required to pay the licence fee by a court, that is, $\partial p/\partial c_2^b < 0$, with $\lim_{c_2^b \to \infty} p > 0$ and $\partial^2 p/\partial {c_2^b}^2 > 0$. Find a subgame-perfect Nash equilibrium of the following game. In the first stage, firm 1 chooses a fixed ϕ. In the second stage, firm 2 chooses both its investment levels. In the third stage, consumers make their purchase decisions and courts enforce the licence fee ϕ with probability $p(c_2^b)$. What changes if firm 1 can set ϕ as a share of r_2 instead of a fixed fee?

19.4 Essay question: Cumulative innovations

In the *Human Genome News Archive Edition* of November 2000, one can read the following: 'The deluge of data and related technologies generated by the Human Genome Project (HGP) and other genomic research presents a broad array of commercial opportunities. Seemingly limitless applications cross boundaries from medicine and food to energy and environmental resources, and predictions are that life sciences may become the largest sector in the U.S. economy. Established companies are scrambling to retool, and many new ventures are seeking a role in the information revolution with DNA at its core. IBM, Compaq, DuPont, and major pharmaceutical companies are among those interested in the potential for targeting and applying genome data. In the genomics corner alone, dozens of small companies have sprung up to sell information, technologies, and services to facilitate basic research into genes and their functions. These new entrepreneurs also offer an abundance of genomic services and applications, including additional databases with DNA sequences from humans, animals, plants, and microbes.'

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

1. Innovations in the genomic field are in essence cumulative. Define two different types of cumulative innovation.
2. For each type of cumulative innovation, describe the generic problem that is likely to arise and discuss how the patent system could be complemented in order to solve these problems.
3. What is your opinion about the fact that ‘dozens of small companies have sprung up to sell information, technologies, and services to facilitate basic research into genes and their functions’? Is this likely to foster, or rather to impede, innovation in the genomic field?

P1: SFK Trim: 246mm × 189mm Top: 10mm Gutter: 19.001mm
CUUK2869-19 CUUK2869/Belleflamme & Peitz 978 1 107 06997 8 February 19, 2015 16:51

References for Part VII

Aguiar, L. and Martens, B. (2013). Digital Music Consumption on the Internet: Evidence from Clickstream Data. Mimeo.

Amir, R., Evstigneev, I. and Wooders, J. (2003). Noncooperative versus cooperative R&D with endogenous spillover rate. *Games and Economic Behavior* 42: 183–207.

Anand, B. and Galetovic, A. (2004). How market smarts can protect property rights. *Harvard Business Review* (December): 73–79.

Anand, B. N. and Khanna, T. (2000). The structure of licensing contracts. *Journal of Industrial Economics* 48: 103–135.

Arora, A., Ceccagnoli, M. and Cohen, W. C. (2003). R&D and the Patent Premium. NBER Working Paper no. 9431.

Arrow, K. (1962). Economic welfare and the allocation of resources for inventions. In Nelson, R. (ed.), *The Rate and Direction of Inventive Activity*. Princeton, NJ: Princeton University Press.

Belleflamme, P. and Peitz, M. (2012). Digital piracy: Theory. In Peitz, M. and Waldfogel, J. (eds), *The Oxford Handbook of the Digital Economy*. Oxford: Oxford University Press.

Belleflamme, P. and Peitz, M. (2014). Digital piracy. In Backhaus, J. (ed.), *Encyclopedia of Law and Economics*. Berlin: Springer (in press).

Belleflamme, P. and Vergari, V. (2011). Incentives to innovate in oligopolies. *Manchester School* 79: 6–28.

Bessen, J. (2004). Holdup and licensing of cumulative innovations with private information. *Economics Letters* 82: 321–326.

Bessen, J. and Hunt, R. M. (2007). An empirical look at software patents. *Journal of Economics & Management Strategy* 16: 157–189.

Bessen, J. and Maskin, E. (2009). Sequential innovation, patents, and imitation. *Rand Journal of Economics* 40: 611–635.

Bester, H. and Petrakis, E. (1993). The incentives for cost reduction in a differentiated industry. *International Journal of Industrial Organization* 11: 519–534.

Bousquet, A., Cremer, H., Ivaldi, M. and Wolkovicz, M. (1998). Risk sharing in licensing, *International Journal of Industrial Organization* 16: 535–554.

Brander, J. and Spencer, B. (1983). Strategic commitment with R&D: The symmetric case. *Bell Journal of Economics* 14: 225–235.

Caillaud, B. (2003). La propriété intellectuelle sur les logiciels. In Conseil d'Analyse Economique (ed.), *Propriété Intellectuelle*. Paris: La Documentation Française.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *Economic Journal* 99: 569–596.

Cohen, W. M., Nelson, R. R. and Walsh, J. P. (2000). Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent or Not. NBER Working Paper no. 7552.

Colombo, S. and Filippini, L. (2014). Patent licensing with Bertrand competitors. *The Manchester School* (forthcoming).

Cournot, A. (1838). *Recherches sur les Principes Mathematiques de la Theorie des Richesses*. Paris: Hachette.

d'Aspremont, C. and Jacquemin, A. (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *American Economic Review* 78: 1133–1137. Erratum: *American Economic Review* 80 (1990): 641–642.

Delbono, F. and Denicolo, V. (1990). R&D investment in a symmetric and homogeneous oligopoly. *International Journal of Industrial Organization* 8: 297–313.

Denicolo, V. (1996). Patent races and optimal patent breadth and length. *Journal of Industrial Economics* 44: 249–265.

Dewatripont, M. and Legros, P. (2013). 'Essential' patents, FRAND royalties and technological standards. *Journal of Industrial Economics* 61: 913–937.

European Commission (2013). *Innovation Union. A Pocket Guide on a Europe 2020 Initiative*. Luxembourg: Publications Office of the European Union.

Galasso, A. and Schankerman, M. (2013). Patents and Cumulative Innovation: Causal Evidence from the Courts. Mimeo.

Galasso, A., Schankerman, M. and Serrano, C. (2013). Trading and enforcing patent rights. *Rand Journal of Economics* 44: 275–312.

Gallini, N. (1992). Patent policy and costly imitation. *Rand Journal of Economics* 23: 52–63.

Geroski, P. (1995). Markets for technology: Knowledge, innovation and appropriability. In Stoneman, P. (ed.), *Handbook of Economics of Innovation and Technological Change*. Oxford: Blackwell.

Gilbert, R. and Newbery, D. (1982). Preemptive patenting and the persistence of monopoly. *American Economic Review* 72: 514–526.

Gilbert, R. and Shapiro, C. (1990). Optimal patent length and breadth. *Rand Journal of Economics* 21: 106–112.

Goyal, S. and Moraga-González, J. L. (2001). R&D networks. *Rand Journal of Economics* 32: 686–707.

Green, J. and Scotchmer, S. (1995). On the division of profit in sequential innovation. *Rand Journal of Economics* 26: 131–146.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 18: 1661–1707.

Grünfeld, L. A. (2003). Meet me halfway but don't rush: Absorptive capacity and strategic R&D investment revisited. *International Journal of Industrial Organization* 21: 1091–1109.

Hagiu, A. and Yoffie, D. (2013). The new patent intermediaries: Platforms, defensive aggregators, and super-aggregators. *Journal of Economic Perspectives* 27: 45–66.

Hall, B. H. and Ziedonis, R. H. (2001). The patent paradox revisited: An empirical study of patenting in the US semiconductor industry, 1979–95. *Rand Journal of Economics* 32: 101–128.

Heller, M. and Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science* 280: 698–701.

Jensen, R. and Thursby, M. (2001). Proofs and prototypes for sale: The licensing of university inventions. *American Economic Review* 91: 240–259.

Johnson, J. P. (2012). Open source software. In Peitz, M. and Waldfogel, J. (eds), *The Oxford Handbook of the Digital Economy*. Oxford: Oxford University Press.

Johnson, W. R. (1985). The economics of copying. *Journal of Political Economy* 93: 158–174.

Kamien, M. and Tauman, Y. (1986). Fees versus royalties and the private value of a patent. *Quarterly Journal of Economics* 101: 471–491.

Kamien, M. I. and Tauman, Y. (2002). Patent licensing: The inside story. *The Manchester School* 70: 7–15.

Kamien, M. I. and Zang, I. (2000). Meet me halfway: Research joint ventures and absorptive capacity. *International Journal of Industrial Organization* 18: 995–1012.

Kamien, M. I., Muller, E. and Zang, I. (1992). Research joint ventures and R&D cartels. *American Economic Review* 82: 1293–1306.

Klemperer, P. (1990). How broad should the scope of a patent be? *Rand Journal of Economics* 21: 113–130.

Koh, B., Murthi, B. P. S. and Raghunathan, S. (2013). Shifting demand: Online music piracy, physical music sales, and digital music sales. *Journal of Organizational Computing and Electronic Commerce* 24: 366–387.

Leahy, D. and Neary, J. P. (1997). Public policy towards R&D in oligopolistic industries. *American Economic Review* 87: 642–662.

Lee, T. and Wilde, L. (1980). Market structure and innovation: A reformulation. *Quarterly Journal of Economics* 94: 429–436.

Lemley, M. A. (2012). Software Patents and the Return of Functional Claiming. Stanford Public Law Working Paper no. 2117302.

Lerner, J. and Tirole, J. (2001). The open source movement: Key research questions. *European Economic Review* 45: 819–826.

Lerner, J. and Tirole, J. (2002). Some simple economics of open source. *Journal of Industrial Economics* 50: 197–234.

Lerner, J. and Tirole, J. (2004). Efficient patent pools. *American Economic Review* 94: 691–711.

Lerner, J. and Tirole, J. (2005a). The scope of open source licensing. *Journal of Law, Economics and Organization* 21: 20–56.

Lerner, J. and Tirole, J. (2005b). The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives* 19: 99–120.

Lévêque, F. and Ménière, Y. (2004). *The Economics of Patents and Copyright*. Berkeley, CA: The Berkeley Electronic Press.

Levin, R. C., Klevorick, A. K., Nelson, R. R. and Winter, S. G. (1987). Appropriating the returns from industrial R&D. *Brookings Papers on Economic Activity*: 783–820.

Liebowitz, S. and Margolis, S. (2004). Seventeen famous economists weigh in on copyright: The role of theory, empirics, and network effects. *bePress Legal Series*, Paper 397.

Llanes, G. and Poblete, J. (2014). Ex ante agreements in standard setting and patent-pool formation. *Journal of Economics and Management Strategy* 23: 50–67.

Loury, G. C. (1979). Market structure and innovation. *Quarterly Journal of Economics* 93: 395–410.

Mansfield, E. (1986). Patents and innovation: An empirical study. *Management Science* 32: 173–181.

Maurer, S. M. and Scotchmer, S. (2004). Procuring knowledge. In Libecap, G. (ed.), *Intellectual Property and Entrepreneurship: Advances in the Study of Entrepreneurship, Innovation and Growth* 15: 1–31. Amsterdam: JAI Press (Elsevier).

Maurer, S. M. and Scotchmer, S. (2006). Open source software: The new intellectual property paradigm. In Hendersshott, T. (ed.), *Handbook of Economics and Information Systems*. Amsterdam: Elsevier.

Menell, P. and Scotchmer, S. (2007). Intellectual property. In Polinsky, A.M. and Shavell, S. (eds), *Handbook of Law & Economics*. Amsterdam: Elsevier.

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

Nordhaus, W. (1972). The optimal life of the patent: Reply. *American Economic Review* 62: 428–431.

Novos, I. E. and Waldman, M. (1984). The effects of increased copyright protection: An analytic approach. *Journal of Political Economy* 92: 236–246.

OECD (2000). Competition and Regulation Issues in the Pharmaceutical Industry. DAFFE/ CLP(2000)29.

Okuno-Fujiwara, M. and Suzumura, K. (1990). Strategic Cost-Reduction Investment and Economic Welfare. Mimeo.

Peitz, M. and Waelbroeck, P. (2006). Why the music industry may gain from free downloading – the role of sampling. *International Journal of Industrial Organization* 24: 907–913.

Qiu, L. D. (1997). On the dynamic efficiency of Bertrand and Cournot equilibria. *Journal of Economic Theory* 75: 213–229.

Reinganum, J. (1983). Uncertain innovation and the persistence of monopoly. *American Economic Review* 73: 741–748.

Rey, P. and Tirole, J. (2013). Cooperation vs. collusion: How essentiality shapes coopetition. IDEI Working Paper 801.

Ridley, D. B., Grabowski, H. G. and Moe, J. L. (2006). Developing drugs for developing countries. *Health Affairs* 25: 313–324.

Schumpeter, J. (1943). *Capitalism, Socialism, and Democracy*. London: Unwin University Books.

Scotchmer, S. (2004). *Innovation and Incentives*. Cambridge, MA: MIT Press.

Shapiro, C. (2001). Navigating the patent thicket: Cross licenses, patent pools, and standard setting. *NBER/Innovation Policy and the Economy* 1: 119–150.

Shavell, S. M. and van Ypersele, T. (2001). Rewards versus intellectual property rights. *Journal of Law and Economics* 44: 525–547.

Simcoe, T. (2012). Standard setting committees: Consensus governance for shared technology platforms. *American Economic Review* 102: 305–336.

Simonite, T. (2013). Copy protection for 3-D printing aims to prevent a piracy plague. *MIT Technology Review*.

Sovinsky, M. and Helland, E. (2012). Do Research Joint Ventures Serve a Collusive Function? Mimeo.

Spence, M. (1984). Cost reduction, competition and industry performance. *Econometrica* 52: 101–112.

Strowel, A. (2005). The Principles and Rationale of Patent Protection. Mimeo.

Takalo, T. (2001). On the optimal patent policy. *Finnish Economic Papers* 14: 33–40.

Tandon, P. (1982). Optimal patents with compulsory licensing. *Journal of Political Economy* 90: 470–486.

Tirole, J. and Lerner, J. (2013). Standard-Essential Patents. Mimeo.

Waldfogel, J. (2012a). Digital piracy: Empirics. In Peitz, M. and Waldfogel, J. (eds), *The Oxford Handbook of the Digital Economy*. Oxford: Oxford University Press.

Waldfogel, J. (2012b). Copyright protection, technological change, and the quality of new products: Evidence from recorded music since Napster. *Journal of Law and Economics* 55: 715–740.

Wang, X. H. and Yang, B. (2004). On technology transfer to an asymmetric Cournot duopoly. *Economics Bulletin* 4: 1–6.

Yi, S.-S. and Shin, H. (2000). Endogenous formation of research coalitions with spillovers. *International Journal of Industrial Organization* 18: 229–256.