Chapter 2: Market Power and Welfare

(PRELIMINARY VERSION: Comments welcome!)

Competition Policy: Theory and Practice
Cambridge University Press, 2003

Massimo Motta
Department of Economics
European University Institute, Florence
and Universitat Pompeu Fabra, Barcelona

7 January 2003

Contents

1 Market Power and Welfare: Introduction

1.1 Overview of the chapter

At the basis of competition policy there is the idea that monopolies are "bad". Section 2 shows that indeed a monopoly causes a static inefficiency: for given technologies, monopoly pricing results in a welfare loss. Further, there generally is an inverse relationship between market power (the ability of firms to set prices above marginal costs), of which monopoly power is the most extreme form, and (static) welfare.

Sections 3 and 4 show that just by looking at allocative inefficiency alone one might actually underestimate the welfare loss from market power. A monopoly (more generally, high market power) might result in productive and dynamic inefficiencies too. Not only does a monopolist charge too high a price, but it might also have too high costs and innovate too little, since - sheltered from competition - it is not pushed to adopt the most efficient technologies and to invest much in R&D.

One might be tempted to conclude that if having one firm (or very few firms) leads to welfare losses, then competition policy should try to increase the number of firms which operate in the industry (for instance subsidising and protecting less successful firms). Section 3 also shows that such a conclusion would not be correct, because keeping less efficient firms artificially alive would distort the allocation of resources and reduce economies of scale, thus reducing welfare. In short: (1) *competition policy is not about maximising the number of firms*, and (2) *competition policy is about defending market competition, not about defending competitors.*

Although there is an inverse relationship between market power and welfare when focusing on a static analysis, it is not clear that the same unambiguous relationship exists when productive and dynamic inefficiencies are considered. In any event, section 4 argues that market power is certainly not *per se* bad. Indeed, the prospect of enjoying some market power (and profits) is the main incentive for firms to invest and innovate. If firms were not able to appropriate the results of their investments - whether they are advertising outlays, capacity, R&D expenditures or other - they would not invest at all, with the result that consumers would not benefit from lower costs, higher quality goods, new product varieties and so on.

Section 5 shows that there often exists a trade-off between ex-post and ex-ante considerations. If ex-post (i.e., for given qualities, technologies, varieties) one would be tempted to eliminate market power so as to reduce prices and increase allocative efficiency, ex-ante such a policy would be detrimental, because it would eliminate the incentives for firms to improve their quality offerings and their technologies. Public policies should guarantee firms some market power, that is appropriability of their R&D and investments, and competition policy should not aim at "destroying" monopolies, or more generally firms' market power, as long as they are established on the basis of legitimate business practices. A firm which enjoys a monopoly after having successfully invested, innovated, introduced new products is a firm which receives a reward for its activities. The expectation that - if doing well - a firm will receive profits is the incentive which pushes such a firm to do well. Any attempt to eliminate market power after a firm has successfully attained it gives the wrong incentive signals to this and all the other firms. Competition policy is therefore not concerned with monopolies *per se*, but rather only with monopolies which distort the competitive process.

The previous arguments suggest that competition policy should not be too interventionist. Some theories go further, and suggest that there exist market mechanisms which prevent a monopolist from exercising market power, thereby reducing further the scope for competition policy. Section 6 discusses these views, and in particular the argument that free entry acts as a restraint to the market power of a monopolist. Although it is true that potential entrants can play a role in disciplining incumbents, monopolists often continue unperturbed to charge high prices even when entry is possible, for a number of reasons, such as the existence of sunk costs, consumer switching costs, and network effects, as well as business practices carried out by incumbents to exclude potential entrants (such practices will be analysed in chapter 7).

Section 7 ends the chapter with some policy conclusions.

2 Allocative efficiency

Section 2.1 defines market power. Section 2.2 shows that market power brings about a welfare loss due to prices being too high: this welfare loss is labelled "allocative inefficiency", and it is highest when market power coincides with monopoly power. Section 2.3 suggests that the welfare loss might be larger when firms engage in unproductive activities in order to grab market power. Other reasons why the welfare loss from market power might be even higher will be analysed in sections 3 and 4.

2.1 Market power: A definition

Market power is a crucial concept in the economics of competition law. It refers to the ability of a firm to raise price above some competitive level - the benchmark price - in a profitable way. Since the lowest possible price a firm can profitably make is the price which equals the marginal cost of production,[1] market power is usually defined as the difference between the prices charged by a firm and its marginal costs of production.[2]

Since market power refers to the ability of firms to charge prices above marginal costs, we expect firms to have some degree of market power in the real world (not least because if they had zero profit they could not cover their fixed costs), with an unchallenged monopolist enjoying the highest possible market power. The issue of how to measure market power in practice is delayed to the next chapter, but note that in many circumstances competition policy will be concerned only with those firms which have "large enough" market power, where "large enough" will be arbitrary to some extent, and might be different according to the particular competition problems being scrutinised.[3]

The concept of "market dominance" which is used in European competition law (see chapter 1) does not have a clear translation in economic terms,[4]

---

[1] In a model where firms sell perfectly identical products, have zero fixed costs, and choose prices (the so-called *Bertrand* model), price is equal to marginal cost at the equilibrium. (See chapter 8 for a brief description.) The same outcome also arises under perfect competition.

[2] An alternative definition might be to define market power not by how far the prices are with respect to marginal costs, but by how close they are to monopoly prices. Conceptually, these definitions are very similar. Operationally, one is not easier to estimate than the other: the former requires an estimation of marginal costs, the latter of the industry's monopoly prices. Since the former definition is more widely used, I shall adopt it.

[3] For instance, competition laws might fix some minimum market power thresholds below which some rules do not apply, under the presumption that firms with little market power can do little damage. Different thresholds might also been adopted for different practices, such as various types of vertical restraints, which have different potential of harming competition. And the market power thresholds used in merger analysis or in cases of abuse of dominant position need not be the same.

[4] It is not clear how the Court's definition "having the ability of behaving independently of customers, suppliers and rivals" could be rationalised in economic terms. Perhaps, it might be interpreted as the possibility of a firm to behave as a monopolist on the residual

but can be interpreted as a situation where a firm has a large degree of market power, which allows it to charge prices which are "close enough" to those that a monopolist would charge.

2.2 The allocative inefficiency of a monopoly

Let us now analyse why market power reduces static welfare. I focus here on allocative efficiency, and assume that technologies (costs) are given, and that the most efficient technology available is used. In what follows, a simple graphical analysis illustrates the main argument: when prices are above marginal costs, this entails higher producer surplus but not higher enough to compensate for the lower consumer surplus caused by higher prices.

A simple graphical analysis Assume for simplicity that there exists a linear market demand, described by the line OO' in Figure 2.1, and a constant returns to scale technology, represented by the line of constant marginal costs $p_c c$. In the most competitive case, our benchmark case,[5] the price is $p_c = c$ and the quantity sold to consumers is equal to q_c. Consider then the extreme case where market power is maximum: the industry is monopolised by a single firm, which charges the monopoly price p_m.[6,7] The equilibrium output would be given by q_m.

INSERT Figure 2.1. The deadweight loss from monopoly

demand curve in a profitable way, but even this interpretation is not without problems.

[5] The most competitive case corresponds to the Bertrand (or perfectly competitive) equilibrium.

[6] The monopoly price is determined as the price at which marginal revenue (the dotted line in the figure) equals marginal cost. To understand this basic equilibrium condition between marginal revenue (MR) and marginal cost (MC), consider the case where, at a certain number of units sold, the monopolist finds that MR>MC. Then it means that by selling an extra unit of output it would obtain a higher profit, since this extra unit would give a higher revenue than cost. Therefore, MR>MC cannot be an equilibrium. By increasing output, the monopolist would go down the curve MR, getting closer to the equilibrium. By the same reasoning, MR<MC would imply that the last unit sold by the monopolist costs more than it brings revenue to it. Therefore, MR<MC cannot be an equilibrium. The monopolist would rather reduce the number of units sold, and by reducing output, MR will increase by getting closer to the equilibrium.

[7] Note that the same price p_m might be set if there were full collusion among several identical firms operating under constant marginal costs.

Recall that welfare is defined as the sum of consumer surplus and producer surplus. Under the most competitive equilibrium, welfare is given by the triangle Op_cS which also corresponds to the consumer surplus (firms do not have any surplus, since profits are equal to zero).[8] Under *monopoly*, welfare is given by the area described by the points Op_cTR, which is itself the sum of producer surplus p_mp_cTR and of consumer surplus Op_mR. The net efficiency loss caused by the monopoly is given by the difference between the areas Op_cS and Op_cTR, namely by the area of the triangle RST, which is the deadweight loss, D, for the economy.[9]

Note also that, relative to monopoly, competition increases net welfare but does not bring about a *Pareto improvement* (i.e., not everybody is better off), since the producer surplus shrinks with respect to the monopoly case. This might be trivial, but it illustrates the main interests behind the different situations. An industry's producers will try to lobby in favour of more protection and less competitive pressure, while consumers and users of the industry products will have an interest in backing proposals of more competition.[10]

The determinants of the deadweight loss A welfare loss occurs not just for the monopoly price but for any price above marginal costs. To see this with the help of the figure above, just redo the same analysis as above by comparing the welfare level attained when $p = c$ with that attained under any arbitrary price $p > c$.

One can also check that the higher the price p the larger the welfare loss caused by market power (the triangle which represents the loss gets bigger as

[8] Consumer surplus is the area that lies between the segment OS in the demand line (which describes the willingness to pay of people who buy the good) and the line p_cc (which describes the price they pay).

[9] The monopolist is not able to appropriate all the consumer surplus which is lost by consumers whose willingness to pay is higher than marginal costs. However, if the monopolist was able to set a different price for each consumer (i.e., if it were able to do perfect price discrimination), then its profit would be equal to the whole area of the triangle Op_cS. Perfect price discrimination is unlikely to happen, since it requires that the monopolist knows perfectly the willingness to pay of each consumer. See also chapter 7.

[10] As discussed in chapter 1, because of the fragmentation of their interests, in practice consumers often do not lobby for more competition, whereas firms are strong advocates for less competition.

p gets close to p_m), suggesting that welfare decreases with market power.[11]

Furthermore, the graphical example also suggests that the deadweight loss caused by monopoly power also depends on the elasticity of market demand. If demand were perfectly elastic (OO' horizontal in the figure), then the monopolist would not be able to set any price above marginal cost (consumers would not buy the good if there was an even slight increase in price). Hence, the deadweight loss would be nil in this case. As market demand elasticity decreases, the ability of the monopolist to charge higher prices rises and the deadweight loss increases.[12]

Finally, the absolute value of the deadweight loss depends on the size of the market. The intercept of demand (the point O) in the figure can be seen as the size of the market. If demand OO' shifted in a parallel way towards the origin, that is if it kept the same slope but had a lower intercept, then the deadweight loss associated with monopoly power would be smaller in absolute terms.

2.3 Rent-seeking activities

The allocative inefficiency identified above might understate the actual negative effect of monopoly. It has been suggested that - when monopolies are allowed - firms might try to use their political influence and lobbying power to keep or increase their monopoly power. In this process, they would use resources which could instead be put in more productive use. Hence, rent-seeking activities enlarge the expected welfare loss from monopoly.

Posner (1975) argues that the social cost of a monopoly should include an area which might be as large as the overall monopoly profit a firm obtains

[11] This result is true with linear demand, but does not necessarily hold for all demand functions. See also the footnote **??** in the technical section **??**.

[12] Let me make two technical notes that can be skipped by readers who are not interested in such details. (1) If demand is perfectly inelastic, the deadweight loss will disappear. Consider the case where consumers are willing to pay up to a price V for a unit of the good: the monopolist will set precisely that price V and - as with perfect price discrimination - appropriate all the surplus of consumers. As a result, market power is the highest but there is no deadweight loss. (2) Tirole (1988: 68 and 88) shows that the *absolute* welfare loss DWL does not necessarily decrease with the elasticity of demand. For example, with a constant elasticity demand function $p = q^{-\varepsilon}$, there exists a non-monotonic relationship between welfare loss from monopoly and elasticity. However, with this demand function the size of the market increases with ε. Tirole shows that the *relative* welfare loss, namely DWL/W (where W is welfare) is indeed decreasing with ε.

(that is the area of the rectangle $p_m p_c TR$ in Figure 2.2).

INSERT Figure 2.2 Possible additional loss from rent-seeking activities

This is because firms would waste resources in activities which do not have any social value in the attempt to maintain or acquire monopoly power. A monopoly, whether private or caused by public regulation, creates rents. Agents would compete to appropriate these rents by bribing officials, by forming lobbying groups, or by resorting to other such activities.[13] Because of competition among such agents, the amount of their expenditure in rent-seeking activities would equal the amount of expected profit from monopoly, which leads to a complete dissipation of the rents. (See exercise 1 for a formalisation.) Posner's (1975: 257) arguments crucially hinges on three assumptions: (1) There exists perfect competition among agents who engage in rent-seeking activities. (2) The rent-seeking "technology" is characterised by constant returns to scale. (3) The costs incurred in obtaining a monopoly do not have any socially valuable by-product. Under these assumptions, it is easy to see that the welfare loss of monopoly should include all the profit as well.

The assumptions made by Posner are questionable.[14] In particular, it might well be that some agents are better than others and very "efficient" in their rent-seeking operations, which means that the dissipated rent, if it exists, does not necessarily coincide with the monopoly profits. Further, the assumption that rent-seeking activities never create socially valuable results is very strong. Advertising outlays, for instance, indicated by Posner as rent-seeking activities, might increase information available to consumers as well as their "perceived" value of the good.[15]

The result that competition in rent-seeking activities among agents might lead to a complete dissipation of rents, which would then be completely wasted for welfare purposes, is certainly a very extreme and provocative one. Nevertheless, the argument makes sense. In their effort to get hold of rents, firms might well use resources which could be put in more productive use

[13] The argument is familiar to those who have studied the political economy of protection. See for instance Krueger (1974).

[14] See Fisher (1985) for a critique of Posner's arguments.

[15] More generally, expenses to increase the probability of obtaining rents will go to agents active elsewhere in the economy. A partial equilibrium framework is not appropriate any longer to compute welfare.

elsewhere. Rent-seeking activities may certainly generate inefficient distortions. The actual amount these distortions cause is an empirical issue whose treatment is beyond the scope of this work.

3 Productive efficiency

We have seen that for given production costs, a monopolist (more generally, a firm enjoying large market power) charges too high a price leading to a welfare loss, called allocative inefficiency. However, there might be an additional welfare loss, called productive inefficiency, if a firm operating under monopoly has a higher cost than if it were operating in more competitive environments.

From an empirical point of view, the evidence is mixed, but points to significant productive efficiency losses, which might even be larger than the more usual allocative inefficiency ones.[16]

This section first shows the possible additional welfare losses which arise if a monopolist chooses inefficient technologies, and then analyses why a monopolist might indeed end up with such inefficient technologies, discussing briefly both the theoretical and the empirical literature on these issues.

3.1 Additional welfare loss from productive inefficiency

Let us consider how a productive inefficiency increases the welfare loss due to monopoly power.[17]

Suppose that while firms operating in a more competitive industry have marginal cost c, a monopolist would operate at a higher costs, say $c' > c$. This will imply that the welfare loss is bigger than the area RTS identified in Figure 2.1. Figure 2.3 illustrates this idea. If the monopoly operates at the higher cost c', then welfare under monopoly is given by the area of $OR'Vp'_c$. At the competitive equilibrium, firms would instead operate at a cost c, and welfare is (like in Figure 2.1) given by the area OSp_c. Therefore, the welfare loss from monopoly is given by the sum of areas $R'ST'$ and $p'_cVT'p_c$. This is clearly a bigger area than the deadweight loss triangle RST

[16] See Scherer and Ross (1990: 668-672). The first work that tried to estimate welfare losses from monopoly (in the US) was Harberger (1954).

[17] See chapter 8.2.1 for a simple example that shows that the more inefficient a monopolist the higher the price it will charge.

which considered only allocative inefficiency: the shaded areas in Figure 2.3 represent the additional welfare loss due to productive inefficiency.

INSERT Figure 2.3 Additional loss from productive inefficiency

3.2 Why is a monopolist less efficient?

The argument is still incomplete though. I have so far argued that *if* a monopolist does not adopt the more efficient technology that firms operating under competition would adopt, then the monopoly entails an additional productive efficiency loss. However, I have not yet explained why one should expect a monopolist to be less efficient. Somehow, it does seem a reasonable claim that a firm that does not face any competitive pressure will not make much effort to use the best available technologies, to improve its products and to innovate, but one should look more carefully at the theoretical and empirical arguments that support this claim.

There are two main arguments which suggest that a monopoly is likely to involve productive inefficiency. First, managers of a monopolistic firm have less incentive to make effort. Second - a Darwinian selection argument - when competition exists, more efficient firms will survive and thrive, whereas less efficient firms will shut down. If a monopoly exists, the market will not operate any selection and an inefficient firm is as likely to survive as an efficient one. I analyse each argument in turn.

3.2.1 Monopoly and managerial slack

The idea that competitive pressure leads a firm to look for the most efficient way to organise its production and reduce its cost is a very old one, which dates at least from Adam Smith and has also been discussed by John Hicks. Leibenstein (1966) introduced the concept of "X-inefficiency" to restate the idea that monopoly power - and the "quiet life" which comes with it - brings in managerial inefficiency.

The idea meets the consensus of economists. However, the idea that a firm might end up operating with inefficient techniques when more efficient ones are available is not self-evident and deserves some explanations. To understand why this might happen, one has to consider that in reality firms are complex organisations, and that decisions regarding the adoption of technologies - and more generally decisions affecting the overall level of efficiency of the firm - are taken by managers who might have objectives other than maximisation of profits. Consider a firm in which there exists a separation between ownership (shareholders) and control (managers). Shareholders care about profits, but managers care about their individual utility, determined by wage, career prospects, as well as the level of effort and time they have to put in the job. The manager might also care about profits (typically, the shareholders will write a contract where his remuneration increases with the firm's profits), but in general he will care about other things too. As a result, when he takes decisions about technologies (or he has to take actions which affect the firm's costs) he might not have the right incentives to adopt the most efficient ones (that is, those which maximise profits).

To understand why a firm might choose inefficient technologies one has therefore to resort to so-called *principal-agent models*, that is models where a principal (say, the owner of the firm) wants to induce an agent (the manager of the firm) to take the actions that maximise her (the principal's) payoff. Within these models, one can study how market structure would affect (via the decisions of the owner) the actions of the manager, and hence the cost of the firm.

Principal-agent models do not unambiguously find that competition reduces managerial slack.[18] A tentative conclusion from this literature is that there might exist a non-monotonic relationship between competition and manager's effort (and hence firm's costs). Increasing competitive pressure in a market where there is a monopolist will lead it to be more efficient, but increasing pressure in a market where there is already a great deal of competition might reduce efficiency. (See section 3.2.2 for a brief account of a model where this relationship emerges.)

Finally, note that managerial slack is an argument which applies not only to productive inefficiency, but also to dynamic inefficiency (quiet life does not push the managers to innovate).

[18] For instance, in Hart (1983), competition increases managers' efforts, but Scharfstein (1988) shows that the result depends on a strong assumption on the managers' preferences, i.e. that they are infinitely risk-averse. If the manager's marginal utility from income is strictly positive, competition would *reduce* managerial effort. Other models dealing with the effect of competition upon managers' effort include Hermalin (1992), Horn, Lang and Lundgren (1994), and Martin (1993).

3.2.2 Empirical evidence on individual firms' productivity

The managerial slack hypothesis is difficult to test empirically, but there exists some evidence that *individual firms' productivity is* higher in more competitive markets.

Nickell (1996) analyses a panel data of around 700 UK manufacturing firms over the period 1972-1986. He finds that the larger market share, a possible proxy for market power, the lower the firms' productivity *levels*.[19] Furthermore, he finds that the stronger competition (measured either by increased number of rival firms or by lower levels of rents) the higher total factor productivity *growth*.[20]

Nickell, Nicolitsas and Dryden (1997), based on similar data and analysis as Nickell (1996), confirm that product market competition improves firms' productivity performance, but also indicate that the role of competition in disciplining managers and improving company performance is lower when firms are already subject to financial pressure or external shareholder control.[21]

Overall, even though the empirical evidence collected so far is far from conclusive, it does seem that competition affects productivity, and therefore that environments where firms enjoy monopoly power would be characterised by lower productive efficiency.[22]

Competition and efficiency: Schmidt's model* In Schmidt's (1997) model, a firm might go bankrupt, and a manager would incur a loss when his firm closes down (a reasonable assumption). A firm's owner has to devise a contract to induce more effort by her manager: since effort (which is not observable by the owner) is costly for the manager, the owner might have to pay some rents to the manager in order to increase his effort. Note that the manager's effort makes the firm more efficient. Therefore, it not only

[19] For a discussion of the possible problems associated with the use of market shares as market power indicator, see Nickell (1996: 733-734).

[20] Hay and Liu (1997) also look at the components of firms' efficiency in nineteen UK sectors and find some evidence that competition does play some role in determining efficiency and market shares of firms.

[21] On the relationship between product market competition, corporate finance and managerial effort (and innovation), see Aghion, Dewatripont and Rey (1998, 1999).

[22] There is also some evidence that trade liberalisation (or trade openness) - an indicator of the degree of competition firms in a country face - results in higher efficiency levels. See Tybout (2000) for a survey.

increases the expected firm's profit but also reduces the probability that the firm will face bankruptcy (a more efficient firm will be more sheltered from negative shocks).[23]

Market competition reduces the firm's profit, and this has two different effects on managerial effort. The first effect, labelled "threat of liquidation effect" is unambiguously positive. When competition rises, the threat of going bankrupt becomes more important (profits are lower), and this pushes the manager to exert higher effort in order to avoid the loss from liquidation. The second effect, though, has an ambiguous sign. When competition rises, profit decreases, and as a consequence the value of inducing a given cost reduction *might* also decrease. If it does decrease, then the owner will not want to pay a rent to the manager in order to induce higher effort and thus obtain a cost reduction. Competition might then reduce effort if this adverse effect on the profitability of a cost reduction is sufficiently strong.[24]

Schmidt identifies sufficient conditions to guarantee that competition increases managerial effort, but these conditions are not mild enough to expect that this is the only realistic outcome. In an extension, he studies the impact of competition upon managers' effort in a Bertrand competition example. Competition is measured here by the number of firms that operate in the market, and effort increases the probability of obtaining an innovation which makes the firm a monopolist (if more than one firm obtains the innovation, they will each receive zero profit). He finds that the highest level of effort is given by the duopoly case,[25] therefore finding a non-monotonic relationship between competition and effort (and innovation) which is reminiscent of some empirical studies.[26]

[23] The assumption that the manager is wealth constrained excludes the possibility of attaining the first best simply by selling the firm to the manager, who would then choose effort in order to maximise profit.

[24] Another way of describing these two effects is as follows. The first effect is a Darwinian one: more competition pushes the manager to do better to fight for survival. The second effect is a Schumpeterian one: more competition diminishes expected future profits and therefore reduces the owner's incentives to induce more effort from the manager.

[25] In the working paper version of the model, Schmidt also analysed the Cournot competition case, finding the same non-monotonic relationship, but with a maximal effort for number of firms in the market higher than two. (It is not the actual number of firms under which effort is the highest, but the existence of a non-monotonic function competition/effort, the conclusion that I want to draw attention on).

[26] For instance, Caves and Barton (1990) look at US data and find that efficiency is the highest for intermediate levels of concentration (as measured by the 5-firm concentration

3.2.3 A Darwinian mechanism: competition selects efficient firms

In an industry where there exist more efficient and less efficient firms, competition will force the inefficient firms to exit and this rationalisation improves welfare in that output will be produced at a lower cost. A related argument is that when competition exists a lot of different projects, products and technologies are possible. The market will then take care that only the best among them will survive and thrive whereas the others will disappear. Obviously, under monopoly this sort of Darwinian argument will not occur.[27] The example in technical section 3.5.1, although extremely simple and unable to reproduce the richness of the "Darwinian" argument, shows that market competition might drive out of the industry the less efficient firm, and that this would be beneficial.

From an empirical point of view, the selection argument predicts that competition will increase *industry productivity* through a process of entry into and exit from the industry. Olley and Pakes (1996) is one of the most careful studies of the reasons behind changes in productivity in a given industry, and gives strong support to the selection effect of competition. They analyse the impact of technological change and deregulation in the industry for telecommunications equipment in the US, in the period 1963-1987. This is a period in which the industry was deeply deregulated. For large part of the Twentieth century, AT&T was given exclusive monopoly in the provision of telecommunications equipment, so that its manufacturing subsidiary, Western Electric, dominated the industry for a long time. Competition in the industry rose gradually, thanks to an antitrust decision in 1968, a regulatory decision in 1975 which allowed the connection of private equipment to the network, and the 1982 Consent Decree which led to the divestiture of AT&T. The creation of seven regional Bell operating companies, free to buy their equipment from any supplier, and which could not produce equipment themselves, effectively completed this deregulation process. As a result, there was considerable entry (and exit) in the sector between 1967 and 1987, due both to domestic and foreign producers.[28] Olley and Pakes first use sophisticated econometric techniques to estimate the parameters of a production

ratio) . Green and Mayes (1991) find a similar result on UK data. See also the discussion in section 4 below.

[27] See Jovanovic (1982) for a formalisation of market selection.

[28] See Olley and Pakes (1996: 1266-1270) for a description of the industry and a statistical account of entry and exit in the period.

function for the industry, and then use these estimates to understand the role of technological change and deregulation in explaining the changes in the distribution of plant-level performance between 1974 and 1987.

They find that it is the larger share in output of the more productive firms which explains the rise in productivity in the industry: "the productivity growth that followed regulatory change seemed to result from the downsizing (frequently the shutdown) of (often older) unproductive plants, and the disproportionate growth of productive establishments (often new entrants)".[29]

The role played by exit and entry in increasing productivity is confirmed also by recent studies on both US and UK data.[30] Disney, Haskel and Heden (2000) also find the exit and entry of *plants* is important in explaining productivity increases. The group of single-establishment firms experienced no productivity growth at the firm-level: the productivity growth of this group came entirely from entry and exit. Likewise, when looking at the group of multi-establishment firms, they find that most of the productivity growth of these firms is due to closing down of the less efficient plants and the opening of new and more efficient plants. In their paper, Disney et al. (2000) also carry out a detailed analysis of the impact of competition upon productivity and confirm that an increase in market power (measured by either market shares or rents) reduces both the level and growth of productivity.[31]

Overall, it seems that empirical works confirm the important role played by competition in selecting the most efficient firms and thus raising productive efficiency. This has also an additional implication for competition policy: if less efficient firms were protected or subsidised, this would prevent market competition from selecting the best firms, which will actually result in higher prices and lower welfare.

[29] Olley and Pakes (1996: 1266).

[30] For the US, see Foster, Haltiwanger and Krizan (1998). In an earlier work, Baily, Hulten and Campbell (1992) had found that more than entry and exit, productivity growth was mainly due to increasing output shares in high-productivity plants and decreasing output shares in low-productivity ones. For the UK, see Barnes and Haskel (2000). Sorting (or "external")effects, i.e. productivity increases due to entry or growth of more efficient plants and exit of less efficient plants is found to account for roughly 30% to 60% of productivity increases, according to different data and studies. The remaining is accounted for by "internal growth", i.e. improved productivity at the plant level.

[31] They also argue that there is a potential selection bias in analyses like in Nickell (1996) which look at productivity growth in a sample of *surviving* firms (see Disney et al., 2000: 22). Correcting for this bias reduces but does not eliminate the impact of competition over productivity.

3.3 Number of firms and welfare

Since market power decreases with the number of firms in the industry, one might be tempted to conclude that the larger the number of firms the higher welfare. This is not the case, however, when firms have to incur (recurrent or set-up) fixed costs. Indeed, the presence of fixed costs - which gives rise to scale economies - implies the existence of a trade-off. On the one hand, a higher number of firms entails more competition in the market and lower prices, which undoubtedly increases consumer surplus (and allocative efficiency). On the other hand, it also entails a duplication of fixed costs, which represents a loss in terms of (static) productive efficiency. The net effect on welfare is a priori ambiguous, as technical section 3.5.2 shows.

This trade-off between allocative and productive efficiency (a larger number of firms increases competition and drives prices downward but at the same time involves a loss in economies of scale) implies that a policy aimed at maximising the number of firms in any given industry would be unsound. If an authority tried not just to guarantee that entry is possible in an industry and that all firms compete on equal grounds, but also tried to use subsidies or other industrial policy instruments to actively promote entry in an industry or to artificially prevent firms from exiting, this would contrast with an economic welfare criterion.[32]

These considerations should act as a warning that competition policy is about *defending competition* and is never about *defending competitors* which are less efficient. In particular, we have seen in chapter 1 that in many cases antitrust laws have tried to defend small firms as such (that is, in the absence of an abuse by large firms, and just because the latter enjoy a more efficient scale of production). This might be a legitimate political or social objective, but it is not justifiable from an economic welfare perspective: "twisting" the market so as to have more firms than market forces left alone would support will decrease welfare.

3.4 Conclusions

This section has considered two arguments suggesting that market power decreases productive efficiency. The first is that competition (which lowers

[32] This conclusion is reinforced when firms are not symmetric (as in section 3.5.1). When there exist firms producing at higher costs, their exit from the industry might carry additional efficiency gains.

market power) pushes managers to make effort and be more productive; the second is that competition selects the more efficient firms, which results in lower market prices. Both theory and evidence seem to give some support to these arguments although with some qualification. One should expect that introducing more competition in a monopolistic industry is welfare improving, but it is possible, in principle, that increasing competition in an industry where a lot of competition already exists *might* not increase welfare further.

In a similar vein, it is not necessarily true that the larger the number of firms in an industry the higher welfare, because of the inefficient duplication of fixed costs.[33]

3.5 Competition and productive efficiency*

In this technical section, I present two simple models. The first illustrates how competition might select the most efficient firms; the second analyses the relationship between number of firms and welfare.

3.5.1 Competition and selection of firms: An example*

To illustrate the possible beneficial effect of competition in selecting the most efficient firms, consider the following example. Assume a homogenous good industry where firms compete in quantities. Firms have different efficiency levels (different technologies): out of a population of n firms, there are nk firms with a high marginal cost technology c_h and $n(1-k)$ more efficient firms whose technology allows them to produce at a marginal cost c_l. Assume that demand is given by $p = 1-Q$, where Q is the aggregate output. $Q = \sum_{i \in L} q_i + \sum_{j \in H} q_j$, where we have denoted with L and H respectively the set of low-cost and high-cost firms. The profit functions are given by $\pi_j = (p(Q) - c_h)q_j$ for any $j \in H$ and by $\pi_i = (p(Q) - c_l)q_i$ for any $i \in L$. One can then write the FOCs $d\pi_j/dq_j = 0$ and $d\pi_i/dq_i = 0$:

[33] Similar comments could be made by keeping fixed the number of firms and looking at the effect of tougher product market competition on welfare in a situation where entry is free, a situation analysed by d'Aspremont and Motta (2000). Since firms have to cover fixed set-up costs of entry, and anticipate lower profits when market competition is tougher, there exists a positive relationship between the toughness of competition and industrial concentration. As a result, markets where competition is the toughest might not be associated with the highest welfare since fewer firms will co-exist at the equilibrium. Yet, a situation where cartels are allowed (that is, a monopolistic outcome) is the worst in terms of welfare. Exercise 4 develops this result in a simple example.

$$-q_j + 1 - \sum_{i \in L} q_i - \sum_{j \in H} q_j - c_h \quad = \quad 0, \tag{1}$$

$$-q_i + 1 - \sum_{i \in L} q_i - \sum_{j \in H} q_j - c_l \quad = \quad 0 \tag{2}$$

Focusing on the symmetric solution (symmetric in the sense that firms of a given type produces the same output at equilibrium), the FOCs can be simplified as follows:

$$q_h = \frac{1 - c_h - (1-k)nq_l}{1 + kn}; \quad q_l = \frac{1 - c_l - knq_h}{1 + (1-k)n}, \tag{3}$$

where q_h and q_l denote the output of a high cost and a low cost firm respectively. The equilibrium solution can then easily be found:

$$q_h^* = \frac{1 - c_h - n(1-k)(c_h - c_l)}{1 + n}; q_l^* = \frac{1 - c_l + nk(c_h - c_l)}{1 + n}. \tag{4}$$

The equilibrium price is given by:

$$p^* = \frac{1 + nkc_h + n(1-k)c_l}{1 + n}. \tag{5}$$

Note that the high cost firms can produce a non-negative output if and only if $c_h \leq [1 + n(1-k)c_l] / [1 + n(1-k)]$. This condition becomes the more stringent the larger n. In other words, if we proxy the degree of competition in the industry with the number of firms which operate in such an industry, then the stronger competition the more likely that the inefficient firms will exit at equilibrium.[34]

Suppose therefore that competition brings the more inefficient firms to exit the market. I now show that, despite the decrease in the number of sellers, price in the industry will decrease. In other words, exit is beneficial because it allows a reallocation of output from inefficient to efficient firms.

To see this, compute the equilibrium if the inefficient high cost firms exit. There will be only the $(1-k)n$ firms with identical low cost technology. The equilibrium per firm quantity and price will be:

[34] Another way to express the idea that competition increases efficiency could be to proxy fiercer competition not by the larger number of firms, but by a switch from Cournot to Bertrand: the high-cost firms would immediately exit the market.

$$q_l^{**} = \frac{1 - c_l}{1 + (1-k)n}; \quad p^{**} = \frac{1 + n(1-k)c_l}{1 + n(1-k)}. \tag{6}$$

One can easily check that $p^* > p^{**}$ if $c_h > [1 + n(1-k)c_l] \,/\, [1 + n(1-k)]$. But the last inequality is precisely the same condition under which the high cost firms exit the industry. Therefore, the exit of the inefficient firms improves welfare through a reduction in the market prices.[35]

3.5.2 Too many firms in the industry*

Consider a homogenous good industry with n firms which are perfectly symmetric. The firms simultaneously decide on the quantity to bring to the market, and have a production cost $C = cq + F$, where c is the marginal cost and F a fixed cost. Market demand is characterised by $p = 1 - Q$, p being the market price and Q the aggregate output.

Each firm i will choose q_i to maximise its profits:

$$\Pi_i = (1 - q_i - \sum_{j \neq i} q_j) q_i - c q_i. \tag{7}$$

The first-order condition is given by:

$$q_i = \left(1 - c - \sum_{j \neq i} q_j \right) /2. \tag{8}$$

By imposing symmetry, so that $q_i = q_j = q^c$, we obtain the equilibrium solution (it is straightforward to check that second-order conditions are satisfied):

$$q^c = \frac{1 - c}{n + 1}. \tag{9}$$

By substitution, one can also obtain the equilibrium price as a function of the number of firms operating in the industry: $p^c = (1 + nc)/(n + 1)$. Note that when the number of firms n increases, the market price decreases

[35] See Vickers (1995) for a more general treatment of the homogenous good model with Cournot competition and asymmetries. In particular, it is not always true that exit brings about a welfare rise, although this occurs under fairly mild assumptions in the case where consumer and producer surplus have the same weight in the welfare function (Vickers' paper also considers the case where consumer surplus has a larger weight than profits).

and the aggregate output $Q^c = nq^c$ increases, so that the consumer surplus unambiguously rises with n. Indeed, $dCS/dn > 0$, since:

$$CS = (1 - p^c)Q^c/2 = \frac{n^2(1-c)^2}{2(n+1)^2}. \tag{10}$$

However, a larger number of firms brings about an inefficient multiplication of fixed costs. This can be easily seen by noticing that the individual firm's profit is:

$$\pi^c = \frac{(1-c)^2}{(n+1)^2} - F. \tag{11}$$

As a result, the producer surplus in the industry is: $PS = [n(1-c)^2] \,/\, [(n+1)^2] - nF$, which is decreasing with n. In particular, note that as n increases and eventually tends to infinity total welfare becomes negative and then tends to $-\infty$.

Therefore, a policy which aims at maximising the number of firms in the industry would be in contrast with an objective of economic efficiency.[36]

4 Dynamic efficiency

So far I have considered static, rather than dynamic, properties of market power: section 2 showed that a monopoly involves an allocative inefficiency, since a firm (for any given technology) charges too high a price; section 3 showed that monopoly involves a productive inefficiency, since a monopolist might not adopt the most efficient technology *available*. This section considers dynamic efficiency, which refers to the extent to which a firm introduces *new* products or processes of production. In other words, whereas section 3 studied whether competition pushes firms to operate at or closer to the current efficient frontier of production, this section considers whether competition pushes them to move the efficient frontier of production faster or more forward.

Section 4.1 shows that a monopolist might indeed have lower incentives to innovate, thus adding dynamic inefficiency to the list of welfare losses created by a monopoly. However, it is not said that a monotonic relationship exists between market power and innovation. Section 4.2 shows that firms are unlikely to make any investment unless they can expect to appropriate them.

[36] Exercise 8 analyses the equilibrium of this game when entry is endogenous.

This implies that the expectation of market power has an important role as it gives firms incentives to R&D, a theme which will be further analysed in section 5.

4.1 The lower incentives to innovate of a monopolist

In what follows, I use a simple example to show that a monopolist might be dynamically inefficient because it has little incentive to adopt new technologies. Suppose that a monopolist has the possibility to adopt a process innovation which allows it to produce at the lower marginal cost c_l rather than at the current cost c_h by paying a fixed cost F. Call Π_l and Π_h respectively the profit made under the (low-cost) new and the (high-cost) old technology. To decide whether to make this innovation or not the monopolist will have to compare the *additional* profit $\Pi_l - \Pi_h$ it will make with the fixed outlays F it has to spend. The new technology is adopted only if $\Pi_l - \Pi_h > F$.

Consider now the same decision about whether to adopt the new technology or not for a firm which operates in a competitive environment. With the current technology entailing marginal cost c_h that all firms have, firms charge $p = c_h$ and make zero profit: $\Pi'_h = 0$. Suppose now that one of them has the chance to adopt the new technology which allows it to operate at cost c_l, whereas all other firms will still be operating with the old technology.[37] By adopting the new technology, the firm will have the possibility to make a profit Π_l.[38] Hence, this firm operating under competition will innovate if $\Pi_l > F$, whereas the monopolistic firm will innovate only if $\Pi_l - \Pi_h > F$, a much stricter condition. Therefore, the monopolist would have lower incentives to innovate given that it would consider the "additional" profit brought by the new technology, while a competitive firm would consider the whole profit brought by it.[39]

This simple example illustrates the idea that monopolistic firms often

[37] This is the case, for instance, when the innovation is perfectly protected by a patent.

[38] For simplicity, I am assuming here that a firm operating under Bertrand (or perfect) competition makes the same profit Π_l as the monopolist by adopting the same low cost technology. This needs not be the case. Technically, I am assuming here that the innovation is "drastic". The technical section 4.3.1 shows that the same qualitative results are obtained if one assumes that the innovation is "non-drastic".

[39] The property that a monopolist gains less from an innovation than a competitive firm is also known as Arrow's *replacement effect*.

have lower incentives to innovate than firms which face rivals.[40,41] The analysis of the formal models in sections 4.3.1 and 4.3.2 confirms that a monopoly is dynamically inefficient, in that it reduces the incentives to innovate and invest. In a nutshell, the intuition behind this result is as follows. Competition pushes firms to invest, in order to improve their competitive position relative to their rivals. The absence of competition (whether because there is only a firm, or because there are several firms but they collude) removes this incentive to innovate, and this in turn means that a monopolist will be less efficient (less innovative) than firms which operate under competition.

This example cannot be generalised to conclude that the more competition exists in a market the more likely that firms will innovate. There is an old debate in the economic literature on the link between monopoly power and innovations. This debate goes as far back as Schumpeter, who suggested that monopoly power encourages research and development efforts.

Both theoretical and empirical research on the link between market structure and innovation is not conclusive, even though a "middle ground" environment where there exists some competition but also high enough market power coming from the innovative activities, might be the most conducive to R&D output.[42] This is precisely the result which emerges from the model analysed in the technical section 4.3.2. In a context where the R&D choices of firms are studied, and the number of firms in the industry is taken as a measure of the degree of competition in the industry, increasing competition when there is already much of it is not necessarily good for welfare. Indeed, firms' incentives to innovate are determined not only by the existence of competition but also by the possibility of appropriating the results of their investment. If competition is too strong, appropriability is reduced, and so the incentive to invest and innovate.

The result is that some intermediate levels of competition might be optimal for innovations and productive efficiency, a result which is found also

[40] Another reason why monopolists might innovate less comes from the managerial slack argument analysed above: quiet life does not put enough pressure to do better onto a monopolist.

[41] Aghion, Dewatripont and Rey (1999) analyse a growth model where entrepreneurs do not aim at maximising profits, but are "conservative": they want to avoid the private cost of adopting new technologies and they will adopt them in so far as they will allow the firm to avoid bankruptcy. The paper shows that in this setting competition disciplines the entrepreneurs and increases adoption of new technologies and growth.

[42] See Scherer and Ross (1990: chapter 17) for a review.

in a completely different theoretical framework (that based on endogenous growth models) and is confirmed in empirical works, that do find an inverted U relationship between competition and innovation.[43]

Note, however, that it would be extremely difficult to use this result for practical policy purposes, for instance to pick up the "right" level of competition. Different assumptions on parameters and on the game being played by the firms would give different results as to which ones should be the optimal levels of competition in the marketplace. Hence, intervening on given industries by reducing the level of competition (whatever the mechanism to obtain it) in order to get closer to some theoretically optimal level of competition would not be justified by any robust theory. The only sound and robust conclusion we can derive from analyses like the one above is that a monopoly (or a cartel) is worse than competitive market structures, because it fails to stimulate dynamic efficiency.[44] Accordingly, steps should be taken to restore competition in markets where there is none.

4.2 Incentives to invest in R&D

Competition stimulates innovations, but so does the expectation of being able to appropriate its investment in R&D through market profits. To show the importance of market power as incentive to innovate, reconsider the simple example of section 4.1, but assume that when there is competition no firm can appropriate the innovation: if one firm adopts the technology all other firms are able to produce at the same cost as well, perhaps because there is no patent protecting the innovating firm, or perhaps because of policies which oblige firms to give away their technology to rivals (*compulsory licensing*). In this case, no firm has an incentive to innovate: diffusion of the technology prevents an innovator from benefiting from it, since after an innovation all firms would charge $p = c_l$ and make zero profit: $\Pi'_h = 0$. The fixed cost F of the innovation could never be recovered, and no innovation will arise under competition.

This simple example is admittedly extreme (see section 4.3.3 below for a slightly more general setting), but its purpose is to draw the reader's attention to the fact that market power (or the expectation to exercise it) has

[43] Aghion et al. (2002) present a growth model that suggests such a relationship and conduct an empirical test (on UK panel data) of this prediction.

[44] Exactly the same interpretation should be applied to the result obtained by d'Aspremont and Motta (2000) and illustrated in exercise 4.

an important role in maintaining the firms' incentives to innovate, invest, introduce new goods, and improve product quality. Eliminating market power cannot be an objective that any public policy should pursue, as I will argue in section 5.

4.3 Models of competition and innovation*

This technical section provides a formalisation of some of the arguments made above.

4.3.1 Monopoly gives fewer incentives to innovate: an example*

Here I present a simple example with deterministic R&D which illustrates the lower incentives to innovate in a monopoly than in a duopoly.

Monopoly Consider first a monopolist operating at a cost c^h and facing a linear demand $Q = 1 - p$. It can adopt a technology which gives lower marginal cost $c^l = c^h - x$ (with $x \in \left[0, c^h\right]$) if it pays a fixed cost F. The monopolist first decides whether to innovate or not, and then decides how much to sell in the market.

The monopolist's profit is given by $\Pi = (p - c^i)(1 - p)$, with $i = l, h$ depending on whether it has low or high cost (that is, if it has innovated or not). From the FOCs $d\Pi/dp = 0$, it is straightforward that the optimal price is given by $p = (1 + c^i)/2$.

Therefore, if it does not innovate, the monopolist will charge $p = (1+c^h)/2$ and earn $\Pi = (1-c^h)^2/4$. If it does innovate, it will charge $p_I = (1+c^h-x)/2$ and earn the gross profit $\Pi_I = (1 - c^h + x)^2/4$, but will have to pay the cost F.

Hence, the monopolist will innovate if:

$$\Delta \equiv \Pi_I - \Pi = \frac{x}{4}\left(x + 2(1 - c^h)\right) \geq F. \tag{12}$$

Duopoly Consider now the duopoly case. The two firms (which sell a homogenous good) face the same market demand $Q = 1 - p$ as above. The firm which charges the lowest price p will face all demand, and the one with the highest price will get no demand. If both firms charge the same price, they will share equally demand.

They play the following game. In the first stage, they simultaneously have to decide whether they want to pay the fixed cost F to adopt the technology which gives lower marginal cost $c^h - x$, or keep the current technology giving marginal cost of production c^h. In the second stage, they simultaneously choose prices. I look for the sub-game perfect Nash equilibrium in pure strategies of this game.

Price stage of the game.

If both firms choose to innovate in the first period, in the second period they will both have the same cost $c^h - x$. This is the standard Bertrand game (see chapter 8.3), and the only solution is that both firms will charge at marginal cost ($p = c^h - x$) and get zero gross profit. If none innovate in the first period, firms will again charge at marginal cost at equilibrium (this time, however, $p = c^h$) and get zero profit.

If only one innovates in the first period, at the price sub-game we have two possible situations: either the innovating firm sets the monopoly price (if this is the lower than the marginal cost of the non-innovating firm); or it sets a price slightly below the marginal cost of the non-innovating firm (if the monopoly price is higher than the rival's cost).[45] The first case refers to a "drastic" innovation, the second to a "non-drastic" innovation. I assume that the innovation is non-drastic, that is the marginal cost reduction is small enough: $x \leq 1 - c^h$. (Exercise 5 examines the "drastic" innovation case.)

Then the innovating firm will charge a price slightly lower than the rival, $c^h - \varepsilon$, and make gross profit $\Pi_{nd} = x(1 - c^h)$.

Innovation stage of the game.

At the first stage of the game firms correctly anticipate the profit they make in the following stage. Firms' payoffs are as follows. If both firms innovate, they will both get $-F$. If none innovate, they will both get 0. If only one firm innovates, the innovating firm will get $\Pi_{nd} - F$ and the other firm 0.

This game has the following solutions: If $\Pi_{nd} \geq F$ the only equilibrium is one where one firm innovates and the other does not. If $\Pi_{nd} < F$, then no firm will innovate at the equilibrium.

Comparison between monopoly and duopoly Either market structure does not affect the adoption of innovation (innovations occur under

[45] Other equilibria in pure strategies are eliminated because in weakly dominated strategies (see chapter 8.3 for a discussion).

both market structures, or do not occur under either), or an equilibrium with innovation occurs under duopoly but not under monopoly. Indeed, it can be easily seen that $\Pi_{nd} \geq \Delta$ for $x \leq 2(1-c^h)$, which is in the domain we are considering. Therefore, for values of fixed cost of innovation such that $\Pi_{nd} \geq F > x\left(x+2\left(1-c^h\right)\right)/4$, there will be an innovation under duopoly but not under monopoly. For values of fixed cost such that $F \leq x\left(x+2\left(1-c^h\right)\right)/4$, an innovation will occur in both market structures. For $F > \Pi_{nd}$, no innovation will be adopted under either market structure.

This analysis confirms that a monopolist's incentive to innovate might indeed be lower than the incentive of firms which face competitors.

4.3.2 R&D and competition**

I present now another simple model, where firms first invest in R&D and then compete in quantities, to show that the degree of market competition has an effect on dynamic efficiency. More particularly, a monopolistic market structure is going to lead to lower R&D investments (and therefore higher production costs) than a market structure where several firms coexist and behave non-cooperatively.

Consider the demand function for a homogeneous good $p = a - Q$, with Q being the sum of individual output. Firm i $(i = 1, ..., n)$ is characterised by marginal costs $c_i = C - x_i$, where x_i is the R&D investment made by firm i. The cost of R&D is given by the function $gx_i^2/2$, where $g > 0$ is a parameter expressing the efficiency of R&D production.[46]

The game is a simple two-stage game. Firms simultaneously invest in R&D at the first stage; they then simultaneously choose quantities in the second stage. As usual, I look for the sub-game perfect Nash equilibrium of the game.

At the *last stage* of the game, it is easily checked that the (Cournot) equilibrium output is given by:

$$q_i^c = \frac{a - c_i + \sum_{j \neq i}(c_j - c_i)}{1+n}, \tag{13}$$

and the profit will be given by $\Pi_i = (q_i^c)^2$. (See chapter 8.3 for an introduction to the Cournot model.)

[46] I therefore look at process innovations, and assume that R&D is a deterministic, rather than stochastic. This simplifies the model without depriving it of interesting insights.

At the *first stage* of the game, each firm will have the following profit function:

$$\pi_i = \left(\frac{a - (C - x_i) + \sum_{j \neq i} (x_i - x_j)}{1 + n} \right)^2 - \frac{g}{2} x_i^2. \tag{14}$$

By taking the first derivative we obtain the FOC for the i-th firm:[47]

$$\frac{d\pi_i}{dx_i} = \frac{2n \left(a - C + x_i + \sum_{j \neq i} (x_i - x_j) \right)}{(1+n)^2} - g x_i = 0. \tag{15}$$

Notice that by taking all the rival firms' R&D levels x_j $(j \neq i)$ as given and equal to $\overline{x}$ ($x_j = x_k = \overline{x}$, for $j, k \neq i, k \neq j$), the marginal effect of R&D on firm i's profit can be written as:

$$\frac{d\pi_i}{dx_i} = \frac{2n\,(a - C + x_i)}{(1+n)^2} + \frac{2n\,(n-1)\,(x_i - \overline{x})}{(1+n)^2} - g x_i = 0. \tag{16}$$

This is helpful because it reveals that there are three different effects at play when deciding the optimal R&D levels. The first one is what we could label *appropriability effect*: the larger demand (or net demand, $a - C$) the stronger the incentive to do R&D. This term decreases with n, and it is highest when $n = 1$, i.e. when there is a monopoly. The second term is a *competition effect*. When there is a monopoly ($n = 1$), this term disappears completely, while it is strictly positive when there is competition: it therefore captures the incentive to innovate determined by the existence of competition. Notice that the competition effect increases with n, but a slower pace as n rises: $2n\,(n-1)/(1+n)^2$ is concave in n, and reaches an asymptote when $n \to \infty$. The third term captures instead the marginal cost of R&D, and the efficiency parameter g is the only parameter affecting it. It is the interplay between these effects which will determine the R&D levels chosen at equilibrium by firms and how the number of firms affects the equilibrium solutions.

If we focus on the symmetric equilibrium $x_i = \overline{x} = x$ we obtain the equilibrium R&D level for each firm:

$$x^c = \frac{2n(a-C)}{g(1+n)^2 - 2n}. \tag{17}$$

[47] The second-order condition requires $g > 2n^2/(1+n)^2$. The stability condition requires instead $\left[d^2\pi_i/dx_i dx_{-i}\right] / \left[d^2\pi_i/dx_i^2\right] = 2n(n-1)/\left[g(1+n)^2 - 2n^2\right] < 1$. This is satisfied for $g > 2n(2n-1)/(1+n)^2$. The latter is clearly a more stringent condition. Note that an increase in n tightens this condition. However, a sufficient condition for stability to be met is that $g > 4$.

From this solution, one can then obtain by substitution the equilibrium levels of quantity, price and profit (omitted for shortness). The total level of R&D in the industry will be given by $R^c = nx^c = 2n^2(a - C)/\left[g(1+n)^2 - 2n\right]$. Note that:

$$\frac{dR}{dn} = \frac{2n(a-C)\left(g(1+n)-n\right)}{\left[g(1+n)^2-2n\right]^2} > 0. \tag{18}$$

The last expression tells us that the more firms in the industry (that is, the more competition in the marketplace) the larger the amount of R&D carried out. However, this does not necessarily mean that welfare increases as n increases. It might be that there exists too much R&D with respect to what is optimal for society (that is, that too many resources are devoted to R&D).

To this purpose, let us first compute welfare under competition. As is standard, we take welfare as $W = CS + PS$. In our case:

$$CS^c = \frac{(a-p^c)nq^c}{2} = \frac{\left[(a-C)ng\left(1+n\right)\right]^2}{2\left[g(1+n)^2-2n\right]}, \tag{19}$$

and total profits are:

$$PS^c = n\pi^c = n\frac{(a-C)^2 g\left(g(1+n)^2-2n^2\right)}{\left[g(1+n)^2-2n\right]^2}. \tag{20}$$

Therefore, welfare will be given by:

$$W^c = \frac{(a-C)^2 ng\left((2+n)g(1+n)^2-4n^2\right)}{2\left[g(1+n)^2-2n\right]^2}, \tag{21}$$

which is a non-monotonic function in n.

INSERT Figure 2.4 Welfare and number of firms

As showed in Figure 2.4, the welfare function $W^c(n)$ increases up to a certain value $n_m(g)$ and then decreases, with a horizontal asymptote at $W = (a-C)^2/2$. In other words, an increase in competition (in the sense of an increase in the number of firms in the market) from low initial levels (when appropriability is high) raises welfare; but when competition is large enough, a further increase in it would decrease welfare, since an increase in the number of firms n does not add much to R&D levels (and hence quantities sold in the market) but it decreases profits.[48] Note that when n increases there will be a larger number of firms which have to pay R&D costs, and this duplication of costs determines a productive inefficiency.[49]

4.3.3 Appropriability and R&D*

Assume a homogenous good duopoly where firms have initial marginal cost levels c. Market demand is $q = a - p$; the cost function is: $C_i = (c - x_i - lx_j)q_i - x_i^2$, where x_i is the R&D investments made by firm i ($i, j = 1, 2$; $i \neq j$). Here R&D takes the form of a process innovation (it reduces production cost) and for simplicity it is deterministic. Also assume that there exist R&D spillovers (or externalities): a fraction $l \in [0, 1]$ of the R&D carried out by a firm i will benefit its rival j (and vice versa). This is a crude but simplifying way to formalise the idea that innovations can be (partly or completely) copied, imitated or reverse-engineered by rival firms.[50] The game is as follows. First, firms have to simultaneously decide their R&D investment levels x_i; then, they compete on prices. The objective here is to compare equilibrium outcomes with and without patents.

Duopoly under price competition Note that one can immediately rule out equilibria where both firms innovate, since at least one firm would not be able to recover the fixed cost of R&D expenditures. At the equilibrium of the whole game, only one firm innovates. In general, there are two possible cases to consider. (1) The innovation is drastic. In this case, the monopolistic price of the innovating firm is lower than the cost of the non-innovating firm. For this to happen, the following condition must hold: $(a+c-x)/2 < c-lx$, or: $x \geq (a - c)/(1 - 2l)$. (2) The innovation is non-drastic. In this case, $x < (a - c)/(1 - 2l)$. I focus here on non-drastic innovations (it can be showed that at the equilibrium of this specific model the optimal level of the innovation corresponds to a non-drastic one).

[48] Boone (2000) finds that increasing competition (as proxied by a switch from Cournot to Bertrand) when firms are asymmetric may lead to lower R&D investments. In industries where the technological gap is larger, the laggard's gain from catching up with the leader may fall with more competition. A similar result is also found in Aghion et al. (2002).

[49] See also section 3.5.2 for another example of this duplication effect.

[50] Assume for simplicity that imitation and reverse engineering are costless.

Let us now look at R&D stage and find the optimal value of x. The profit function of the innovating firm will be given by: $\pi_{nd} = x(1-l)(a-c+lx)-x^2$. Note that $x(1-l)$ is the profit margin of the innovating firm: the higher the spillover l the lower its market power.

From solving $d\pi_{nd}/dx = 0$ it follows that the function π_{nd} has a maximum at $x^*(l) = (1-l)(a-c)/\left[2-2l(1-l)\right]$, which is the equilibrium level of R&D. Note that the higher the spillover l the lower the R&D made at equilibrium.

Patent (no spillovers) Imagine now that a patent is awarded to the firm which makes an innovation. As before, we know that the two firms would compete à la Bertrand: at the equilibrium in pure strategies only one of them will do a positive amount of R&D. The treatment is the same as in the previous case, but for the particular value where $l = 0$, i.e., no spillovers could arise (the patent is protecting the innovating firm from spillovers).

The optimal level of R&D under patents will therefore be $x^p = x^*(l = 0) = (a-c)/2$. Clearly, $x^p > x^*(l)$ for all $l > 0$: there is always more R&D under patents.

Conclusions In this simple model, a patent improves welfare. Because of spillovers, R&D is a public good. Since firms are not able to appropriate their R&D efforts, they will invest less in R&D than what society would find it optimal. The patent removes the negative externality given by the spillovers and restores the incentive to make R&D.

It should be noted, however, that this model is rather extreme in that it suggests that patents might be optimal whatever the level of spillovers. This needs not be the case under more general models. Here there are homogenous goods and price competition, so competition is very fierce and has a very strong impact on the returns of R&D. If we assumed weaker competition (for instance, by allowing for some product differentiation) then we might obtain that patents increase R&D levels when spillovers are large enough.[51]

A final qualification is as follows. This simple model does not capture another important feature of patents, that is the fact that a firm is granted a patent only if it discloses completely the technology to other firms. The purpose of this rule is that all rival firms are able (at least in principle) to

[51] Under Cournot competition and homogenous goods, market competition is "too weak" and as a result spillovers do not dissipate much the incentive to innovate (at least at the symmetric equilibrium). As a result, a patent might not be better than allowing spillovers.

share perfectly the technology with the innovating firm and use it as soon as the period of patent protection expires.

5 Public policies and incentives to innovate

From the previous sections it emerges that market power reduces allocative efficiency, but also that a clear-cut relationship between market power on one side and productive and dynamic efficiency on the other side is more difficult to establish. Therefore, there would be no justification for the elimination of market power as an objective of competition policy.

More importantly, the very existence of some market power helps competition. It is precisely the prospect of enjoying some market power (i.e., of making profit) that pushes firms to use more efficient technologies, improve their product quality or introducing new product varieties. If antitrust agencies tried to eliminate or reduce market power whenever it appears, this would have the detrimental effect of eliminating firms' incentives to innovate.

This section stresses the importance of the incentives to innovate (and more generally, to invest), and argues that public policies should maintain such incentives.

5.1 Ex-ante v. ex-post: property rights protection

The trade-off between ex-ante efficiency (one wants to preserve the firms' incentives to innovate) and ex-post efficiency (once firms have innovated it would be better if all the firms in the economy had access to the innovation) is at the core of public policies towards investments and innovations. A government faces a time-consistency problem here. It would be better off by promising firms that they can fully benefit from their R&D results, only to renege on the promise once such R&D results are obtained. But firms would anticipate that the government has such an incentive to renege on the promise to make maximal diffusion of the innovation later on. As a consequence, firms would not innovate at all, resulting in the worst possible solution for welfare. The government is therefore hurt by the possibility to renege on the promise, that is by its lack of commitment. It would be better off by tying its hands and commit to keep its promise. This way, firms would be ensured that they will be able to enjoy the rewards from their R&D.

Patent laws (and other intellectual property rights) are a way for governments to commit not to expropriate an innovating firm ex-post. A firm knows that for a certain period of time it will be able to exploit fully its R&D results.

The trade-off between the necessity of granting firms the appropriability of their innovations and the desire that the benefit of such innovations spread to other firms and to consumers has given rise to a vast literature on the optimal design of patents.[52] This literature aims at identifying the optimal breadth and length of patents. In theoretical terms, the problem is the following. One does not want to give too broad a protection to an innovation, otherwise rival firms would be discouraged to introduce innovations which are only vaguely related to the patented innovation. On the other hand, the protection cannot be too narrow, otherwise a rival firm might make any small and artificial incremental innovation and get away with the infringement of the patent. The same is true for the length of the patent. A too long period of protection makes it impossible for rival firms to challenge the incumbent innovator with new discoveries; but too short a period of protection would not give enough appropriability for innovators. In practical terms, though, it is difficult to go beyond the recognition of such a trade-off and it is very hard to identify a breadth and length of protection which is not arbitrary.

Besides patent laws, there are other laws which help a government solve the commitment problems in a relative straightforward way. Copyright and trademark laws ensure that a successful product or brand cannot be used by another firm unless an agreement with the firm which owns the rights exists. Trade secret laws aim - among other things - at making sure that employees who have been informed of business secrets in a firm cannot bring them to a rival firm which hires them. Non-competition clauses also help firms keep their know-how intact, as they limit the freedom of workers to move to rival firms.[53]

More generally the issue involves not only intellectual property rights but property rights *tout court*. Imagine that a firm invests in a certain plant or in a certain physical asset that later turns out to be a crucial competitive advantage for all other firms in the industry. A competition agency might then be tempted to give access to such a crucial asset to all firms: again if this was possible firms would not make much effort to invest in the first place.

[52] See Tirole (1988: chapter 11) for references.

[53] See Motta and Rønde (2002) for a formalisation of such clauses.

Property laws are there to solve this commitment problem by ensuring that expropriation will not occur.

A similar problem occurs with respect to pricing. Suppose that a firm has introduced a new product or created such a strong brand image of a given product that consumers would love to buy it even at extremely high prices. Clearly, the high price reduces allocative efficiency, and a government might then be tempted to intervene and impose price ceilings to increase surplus. It is therefore important to have a clear legal framework which guarantees firms against such a possibility. We elaborate on the last two points in the following sections.

5.2 Essential facilities

Any input which is deemed necessary for all industry participants to operate in a given industry and which is not easily duplicated might be seen as an *essential facility* (see also chapter 6). There are many examples that might satisfy this very loose definition of essential inputs. In the airline industry, slots in an airport; for maritime transportations, a port's installations; in fixed telephony, the local loop which links each home's telephone with the network; for electrical power generation, the transmission and distribution network of electricity; for the production of pharmaceuticals, a certain chemical component; and so on.

These are theoretical examples, in the sense that understanding whether a given input truly is an essential facility is a complex issue. How indispensable is a certain input? Very often, an input gives its owner a competitive advantage over rivals which have inferior inputs. Obviously, one cannot call "essential" an input which just gives a minor competitive advantage to a firm. On the other hand, there might be alternative inputs that are such poor substitutes that they would not allow rivals to compete at all. It might therefore be not easy to establish where a certain input begins to be so superior as to be judged "necessary".

How easy is it to duplicate an input? Clearly, if reproducing the input was easy, cheap and practical, then it would be difficult to say that it is an essential input. But how costly and difficult should reproduction be in order to conclude that the input is essential?

Imagine for instance that a shipping company X integrates backwards and builds new port installations in a certain town A, located in the "home" country. Given the location, using this port's infrastructure gives firm X a great advantage in serving the maritime route from the home country to a certain other foreign country. Company Y now requests use of the port and firm X denies it (*refusal to supply*). Then, firm Y complains with the competition authorities that it should also have access (possibly even at a price, provided it be fair) to town A's port installations.

Competition authorities in different countries have often been too ready to accept allegations like that made by firm Y, and too easily embraced the so-called *essential facilities doctrine*, thereby deciding that the owner of the input had engaged in an illegal practice and was obliged to make the facility available to competitors.[54]

But there are a number of considerations which should be properly analysed before granting rivals access to a facility owned by a firm. In the port example, for instance, it should be seen to which extent transport from ports other than port A is really such a poor substitute for the maritime route to the foreign country. Are really all the other ports so far away? Are their facilities so inferior? Second, supposing that there are no other existing ports which might provide a substitute route for transport to the foreign country, is it feasible for company Y to reproduce a similar investment and build (or improve) port facilities in another nearby town, say B? Notice that here it would not be enough to answer that it would be too expensive. It should be argued that there are no other towns in the area which are served by train and/or highways so that shipping would be impossible; or that the government would not authorise the building of other ports in a close area because of environmental of other reasons, and so forth. Third, it should also be checked that by letting rivals access the infrastructure the owner of the facility would not find it more costly to produce. If, for instance, there is no spare capacity, then it would be more difficult to argue for access.

If all the previous tests were favourable to the firm requesting use of the input, the possibility of using access pricing to cope with this issue could be considered. However, there is another argument which suggests caution

[54]In a stream of cases related to the transport sector, the EC has been too ready to find "essential" facilities and abuses of dominant positions on the side of the firms that owned such facilities. Most of these cases were related to port infrastructures: *B&I Line v. Sealink Harbours* and *Sea Containers v. Stena Sealink* (both concerning the port of Holyhead in Wales), *Stena v. DSB* (port of Rodby in Denmark), and *Irish Continental Group v. CCI Morlaix* (port of Roscoff in Brittany). However, the subsequent decision by the ECJ in the *Bronner* case has set a higher standard for the application of the essential facilities doctrine, putting a break on the interventionist approach of the Commission.

before granting access to rivals too generously. In our example, the new port installation has been the result of a deliberate (and presumably costly) investment by a firm. Obliging it to share its facilities with rivals would be an infringement of its property rights. More importantly, it would have the effect of discouraging similar investments elsewhere, as the prospect of being expropriated of own investments will make firms refrain from introducing new inputs and facilities in the first place.

The last observation is crucial. There is an important difference between firms which have invested and firms which have obtained the right of using a certain facility without having borne the risk of its creation or having paid for it. In the above example, one should not consider in the same way the case where firm X has built itself (bearing the risk and cost of the investment) the port installations and the case where the investment has been made by the state and the firm has received the monopoly rights to use them.

Take for instance the case - common in Europe - where the "flag carriers" had received by their government, at the time where they were the national monopolists, the right of using most of the landing and take-off slots at domestic airports. Availability of slots in a given airport is an essential input for many routes (alternative, less crowded, airports might be too far away for most consumers, for instance), and once deregulation occurs and entry in the sector is "free", it would be hard to argue that the slots should be kept by the former monopolist. In this case, even if the monopolist holds a contract which gives it exclusive rights for the slots, competition would require letting competitors have access to some of them: there is no issue of protecting the firm's investments here.

Similarly, in some instances firms might obtain intellectual property rights protection over an input without having made any innovation which is worth being protected. In my view, this has been the case in two important EU cases, *Magill* and *IMS*. In *Magill*, a small company (Magill) was accused of infringing upon the copyright of three TV chains because it had offered a weekly TV Guide that included the programmes of all the three broadcasters. Before Magill, the only products available were each broadcaster's weekly TV Guide (that did not include the other TV's programmes), or the listing of all programmes by newspapers, but only for the current day (such information was freely supplied to newspapers by the TV chains). It is doubtful that the weekly listing of programmes by each broadcaster is an innovation which deserves copyright protection (especially if considered that broadcasters spread the relevant information).

In *IMS*, a German firm (IMS) collected data about pharmaceutical products sales in Germany. To do so, it had divided the country into zones, thus creating a reference map that firms in the industry had helped to draw and were using. When a new competitor tried to offer the same service, it had to organise sales data according to the same map (firms refuse an alternative way to organise data because accustomed to the old system), over which IMS claimed copyright. In this case as well, it is difficult to see the map as an innovation worth being protected.

5.3 Price controls and structural remedies

While regulation generally provides regulatory authorities with some sort of control on the prices charged by the firms in the sector, this is not the case for competition policy authorities, whose intervention in the marketplace is assumed to be occasional. Use of price controls and price caps in general contrasts with competition policy. However, many competition laws, such as for instance art. 82 of the EU Treaty, allow authorities to intervene if prices set by a dominant firm are "too high" (see chapter 1.4). These are very dangerous provisions, for at least two reasons. First, deciding if a price is too high or not involves a high degree of arbitrariness. Second, and more important, even if it was established that a firm is charging too high a price (whatever this means), why should it be punished for it? Imagine that the firm is charging the monopoly price (which is the highest profitable price a firm would set). Such a dominant position in the market might come from different sources.

First, it might come from past illegal behaviour, such as predation. If it did, then the competition authority should open a proceeding directly for the infringement of some competition laws, rather than for high prices.

Second, it might be due to present legal barriers to entry. In this case the sector should be subject to regulation, since market forces are not free to operate. If legal barriers to entry have already been removed, though, an industry regulator should operate for a transition period, because free entry does not by itself and immediately eliminates the power of the incumbent (see section 6 below). If for some reason there are no sector regulators and the competition authorities are competent, then it makes sense that they intervene, either by imposing lower prices or, in exceptional circumstances, by the use of structural remedies, that is measures that modify the property rights of a firm, for instance by imposing the disposal of certain assets (patents, brands, plants) or - even more dramatically - the breaking of a firm into several units. This measure should be taken only in really extra-ordinary cases, and should be adopted only if all other measures have failed, if it is clear that incumbents will never be challenged by entrants, and if the dominant position has been acquired not on the merits but because of a legal monopoly.[55]

Third, the firm might have acquired its market power through investments, innovations, and advertising (and maybe even a good share of business luck). If this is the case, then there is no reason to punish it for it. The firm has the right to set high prices since these are the reward for its investments. Intervening by imposing lower prices would be tantamount to depriving it of its risky investments, and discourage this and other firm to invest in the future.

In such a case, the competition authority should just be vigilant, and intervene only if the dominant firm engages in exclusionary practices aimed at preserving its market position. If predation does not occur (and that would be the case in most situations) then either the firm keeps on investing to keep its position, or other firms will sooner or later challenge it and prices will decrease over time.

For the same reasons, targeting firms because they are "too profitable" is wrong. High prices and high profits might be an indication that it is worth looking at the industry (for instance, there might be collusion, or some other infringement of competition law), but certainly they cannot justify by themselves intervention by competition authorities.

5.4 Internal vs. external growth

The view proposed above is one where firms should enjoy market power if it is the result of their investment. Market power, and profit, are the legitimate reward for the fact that a given firm has been more successful than others (whatever the reason, provided not illegal). Call this situation where a firm acquires market power and grows on the merits "internal growth". Contrast it now with a situation - call it "external growth" - where a firm grows and acquires market power not because of its investments but simply because it takes other firms over (or merges with them). In this case, market power is

[55] The only case of break-up in the recent anti-trust history is that of Bell, which was broken up in regional companies (the so-called Baby-Bells). This had indeed been the legal monopolist in the telecom industry for a long time.

not the legitimate reward of some risky activities of the firm, but it is the product of the direct elimination of competitors via compensation (that is, at the price of the takeover). Conceptually, this is quite a different thing, and competition authorities should take care to avoid that this occurs. Chapter 5 will show that a merger between small firms is unlikely to create anti-competitive problems, whereas a merger between firms holding large market shares should be allowed only if it creates substantial, verifiable, merger-specific efficiency gains. That is, if there are strong reasons to believe that by combining the two firms together the result is a much more efficient firm, able to produce at lower cost and sell at lower prices.

6 Monopoly: will the market fix it all?

It has been suggested that market mechanisms prevent even a monopolist from exercising market power. One such argument says that a durable good monopolist is unable to keep prices high because consumers anticipate that it will reduce prices in the future. Another says that if there is free entry, this will prevent the monopolist from setting high prices, as they would trigger entry. Taken at face value, these arguments - derived from the *Coase conjecture* and the *contestable markets theory* - would imply that one should not worry about market power, at least when durable good sellers are involved and when there exists free entry in an industry. In this section, I present and discuss both arguments, to conclude that - although they offer interesting and valuable insights - the extreme versions of such theories are unlikely to hold. I will also mention a number of reasons why market forces alone are unlikely to reduce market power in a number of cases (if sunk costs are important, if consumers have switching costs, if there are network externalities, if a monopolist can engage in anti-competitive practices). Unfortunately, even when entry is in principle free, reasons to worry about monopolies still exist.

6.1 Durable good monopolist

Coase (1972) suggested that the producer of a durable good might price at marginal cost even if it is a monopolist.[56] To gain some insight for this result, suppose that a producer is facing two groups of potential clients who have

[56]Durable goods are goods that last for a certain period of time such as cars, computers, photocopiers, dish-washers and so on.

a different valuation for the durable good it sells. Presumably, the producer will first try to charge a high price and sell to consumers who have a high valuation for the good. In the period after these consumers have bought the good (being a durable good they will not buy again), it would then decrease the price to be able to sell to the low valuation consumers as well. However, it is reasonable to expect that the high valuation consumers understand that the producer will decrease the price in the future. Unless they incur a high cost from delaying the purchase by one period, they will abstain from buying until the price for the good is decreased. As a result, the monopolist will not be able to sell at a high price in the first period.

Suppose now that there are not only two groups of consumers but an extremely large number of them, and that they have valuations for the same durable good that are between the marginal cost and the monopoly price. (Suppose too that each period is very short, so that price adjustments occur continuously.) The monopolist would then have an incentive in each period to reduce the price to sell to those who have not previously bought. Since each consumer will expect that the monopolist will eventually reduce the price to marginal cost to sell to the lowest valuation group of consumers, each consumer will postpone buying until price is equal to marginal cost. We might therefore arrive at the paradoxical result (conjectured by Coase) that the monopolist will lose all its possible market power because consumers anticipate it will reduce prices down to marginal cost in the future.

The Coase conjecture has been formally proved, but it relies on such "heroic" assumptions that it should not be taken literally.[57] However, the insight behind the result is valid, important and applies to several situations (see for instance chapter 6, where this principle is applied to understand the logic and effects of vertical restraints). The crucial issue is that the monopolist is hurt by its *flexibility* to change prices in future periods. If it could commit not to reduce prices in a credible way, then it would be better off. Suppose for instance that it could publicly announce that it charges the monopoly price for its durable good and that it will not change the price as long as the good is sold. Suppose also that the monopolist writes a contract (a so-called *most-favoured nation* clause) specifying that if it ever did decrease the price all consumers who have paid a higher price are entitled

[57]Notably, the assumptions are that both monopolist and consumers live forever, that the good also has infinite durability, that consumers have unit demand, and that the period between successive price adjustments tends to zero. See Tirole (1988: chapter 1.5) for references, a formalisation and discussion.

to the reimbursement of the difference between the original price they paid and the lower price. Then, this contract (enforceable in law courts) would allow it to credibly commit to monopoly pricing and the monopolist would be able to restore all its market power.

Note, however, that the pure announcement of intending not to reduce the price, without any explicit and enforceable commitment, would not be enough to solve its problem. After the first group of consumers have bought the good, the monopolist would have an incentive to renege on the promise and would find it profitable to decrease price.[58]

In conclusion, price flexibility might indeed hurt a seller of durable goods, as well as of other goods which have similar characteristics. But what the example above tells us is that the monopolist might also find a way to solve the commitment problem.

Some instruments solve (or alleviate) the commitment problem As seen above, one possible way for the monopolist to evade the Coase problem is to resort to contractual clauses (such as most-favoured nation clauses) that make its commitment not to lower prices credible.[59]

Another possibility is that, instead of selling the durable good, the producer might rent it (or lease it) to customers. In this way, it commits itself not to decrease price in the future, since a decrease would hurt it by reducing the value of a good it owns. (Exercise 14 formalises this idea and shows that leasing gives higher profit than selling to a durable good monopolist.) One should be aware, though, that leasing is not always feasible or without problems, due for instance to moral hazard problems (consumers who do not own the good might not take care of it and damage it) and the associated monitoring costs.

[58] The problem of inter-temporal credibility and commitment is one which can be found in several circumstances and across fields. Consider for instance a government that wants to stimulate foreign direct investment into the country. It would have an incentive to promise an exemption from corporate taxes. But once firms have established themselves in the country, the government might be tempted to make foreign affiliates to pay high taxes. Similar issues might apply to monetary and macroeconomic policies as well.

[59] Of course, the clauses themselves must be credible. If for instance the price is not publicly observable (when the good sold is an intermediate good and the buyers are firms this is generally the case) then a most-favored-nation clause might be of little use, since a buyer would not easily know prices charged to other buyers later. In such a case, the monopolist will have to find another solution to the commitment problem.

Yet another possibility is that if there is not one, but a stream of products that the monopolist will bring to the market, it might be able to alleviate the commitment problem through establishing a reputation of not flooding the market of a good after the introductory periods.

Conclusions Although theoretically a durable good monopolist might be unable to charge prices over marginal costs, in practice it is unlikely that its market power will disappear because of the commitment problem. If consumers incur costs when delaying their purchases, potential buyers increase over time, or menu costs reduce the monopolist's flexibility in changing prices over time, the monopolist will be able to set prices close to the monopoly level. Above all, a durable good monopolist will have several instruments available, such as leasing, reputation, or various contractual clauses, to solve (or alleviate) the commitment problem and exercise all its market power.

6.2 Contestable markets

It is sometimes argued that monopoly power is likely to be a temporary situation, since the existence of profits would attract the entry of new firms and erode market power. Even if at a given point in time the static deadweight loss was large, from a dynamic point of view the overall expected loss from monopoly power would be much smaller, because of new entrants.[60] If this was the case, there would be little scope for competition policy, since market forces would re-establish a more favourable social outcome without the need of any anti-trust actions.

It is therefore important to analyse whether free entry is likely to decrease market concentration, or in any case to decrease the welfare loss which is due to monopoly power. A useful starting point to discuss this issue is the so-called "contestable markets theory", proposed by Baumol, Panzar and Willig (1982).

A simple illustration of the argument proposed by these authors might go as follows. Consider an industry which produces a homogeneous good by means of a technology which is equally accessible by both an incumbent monopolist and a potential entrant. In particular, to produce the good a firm incurs a fixed cost F plus a variable cost cq. Assume also that the market is

[60] See for instance Schmalensee (1982) and Ordover (1990).

large enough for the monopolist to earn enough profits to recover the fixed cost F.

One then obtains the surprising result that the incumbent firm will not charge the monopoly price, but the price which is just enough to cover its average cost ($p_a = c + F/q$). The proof of this result is straightforward and can be obtained by *a contrario* reasoning, as follows. First, if the monopolist charged a price above average cost, it would obtains positive profits; attracted by these profits a new firm would enter, charge a price slightly lower than the price set by the incumbent, steal all the market to the latter, and earn positive profits. Hence, price above average cost cannot be an equilibrium.

Second, at an equilibrium the incumbent cannot set a price below average costs, because it would not be able to cover its fixed costs and would accumulate losses.

Therefore, price equal to average cost must be the only possible equilibrium outcome under free entry.

The result that a monopolist would not charge a price exceeding average cost is striking. If this outcome was robust and based on reasonable assumptions, its implications for competition policy would be extremely strong. Since it is the very presence of potential entrants in the industry which disciplines the monopolist, any market where entry is not forbidden would attain the socially efficient output (at least when government subsidies are not admitted). Also, since the monopolist would continue to be the only firm operating in the sector, duplication of fixed costs is avoided, and technological efficiency is attained. Indeed, at the equilibrium one would not observe any entry in the market, because no firms would start operating in the industry when the monopolist makes zero profit. Therefore, even the finding that in a given market there exists a persistent monopolist and no entrants over the years would be perfectly compatible with a socially efficient situation. Antitrust authorities would have no reason to intervene in such a world.

Unfortunately, there are two problems in the reasoning above. The first problem comes with the assumption that the monopolist is not able to change its price when it faces new competition. The argument above hinges crucially on the hypothesis that the incumbent sticks to the pre-entry price for at least some time after observing actual entry by another firm. In other words, there would be a higher flexibility in an entry decision by a potential producer than in the price decision of the incumbent. This is unrealistic. The degree of commitment of a firm to its pricing decision is generally low and one would not expect a firm to have such rigid prices that it cannot revise them downwards while it observes a new firm first preparing entry, and then actually entering, its market. In turn, the potential entrant knows very well that the incumbent is able to modify its price and it would rationally expect the price after entry to be lower than that before entry. If there is no room for two firms at more competitive prices, no entry will occur in the industry.

The second problem lies in the nature of the fixed costs F. The theory describes a situation where the potential entrant observes the existence of profit in a given industry, quickly converts part of its existing capacity (or establishes a new plant) to produce the good demanded in that market and sells it. It earns positive profits for some time and then, when the monopolist finally reacts and lowers its price to match the entrant's, would exit the industry and reconvert its capacity to the old production without additional cost (or sell the plant to somebody else, at the same value as the purchase value).

This story depends on the fact that the entrant does not incur any *sunk* cost to start production in the new sector. A sunk cost is a cost which is not recoverable. Generally, setting up new capacity or a new plant involves some degree of fixed sunk costs. To establish new production requires some dedicated, specific investments which cannot be easily recovered, at least not totally, after ceasing production. As an example, an economist which decides to work in industrial organisation has to make an initial investment and devote part of her time to get familiar to the work done in that area. When "recycling" herself into another field (say, macroeconomics), a part of her investment will be lost, since the other field will use different methods. Likewise, to start production in whatever sector involves some fixed set-up costs which are not perfectly recoverable at the moment of leaving the industry. The theory of contestable market requires that there are zero sunk costs. If the entrant had to commit resources in order to produce in the industry, then the "hit-and-run" strategy that this theory proposes would not be possible any longer.

Although its extreme result (that a monopolist would charge at average cost) is unwarranted in most cases, the contestable market theory has had the merit of underlining the role played by potential entry in constraining the market power of incumbents. It is now commonly accepted that a firm is unlikely to exercise such market power if it faces potential rivals that could rapidly and cheaply enter the industry. In many markets, firms have to bear large fixed sunk costs and their investment projects are long and risky. But when this is not the case, one should expect potential entrants to limit significantly prices of established firms. This important insight is crucial for merger policy, as discussed in chapter 5.

6.3 Monopoly and free entry

The contestable markets theory invites to think more seriously on the relationship between monopoly (more generally, market power) and entry. Recent oligopoly models show that free entry is not enough to guarantee that market power in the industry decreases. In this section, I first show that a highly concentrated industry structure can arise under free entry even when firms are ex-ante identical; then, that there are a number of reasons why it might be difficult for entrants to challenge incumbent firms, even if there are no legal barriers to entry. Of course, monopolists might also resort to exclusionary practices to hinder entry.

Monopolies might exist under free entry Consider a situation where many firms might decide to enter an industry. They are all endowed with the same technology and knowledge, they have to incur the same fixed sunk costs if they enter, and they decide simultaneously. They all know that if one firm only enters the market, it will get a large profit, which outweighs the fixed costs. But they also know that, if more than one firm enters, competition will be so fierce (this is a crucial assumption) that even duopoly profits will not be enough to cover fixed costs. Then, as technical section 6.3.1 shows, a monopoly will arise at equilibrium. The expectation of intense product market competition will prevent more than one firm from operating in the industry.[61]

Persistence of concentration when (endogenous) sunk costs exists The previous result is certainly an extreme one. However, a more general and robust result exists. It is called "finiteness property", and states that in markets where consumers value quality of the products, industrial concentration will not fall below a certain threshold even if market size grew arbitrarily large. Intuitively, this is because as demand increases, firms enter an escalation of R&D or advertising outlays (i.e., endogenous sunk costs) to raise the quality of their products. In turn, only a limited number of firms

[61] The milder product market competition the higher the number of firms that will enter at equilibrium. See d'Aspremont and Motta (2000) and Sutton (1991).

can profitably sustain these large endogenous sunk costs. This property is briefly discussed in technical section 6.3.1.

6.3.1 Concentration under free entry*

A simple example* Consider a homogenous product which can be produced by two different firms which are perfectly identical (they possess the same technology and are equally efficient) and analyse the following game. In the first stage of the game the two firms have to decide whether to enter the industry or not. If they do, they have to incur a fixed set-up cost f which is necessary to start production. Once made the investment, the cost f is sunk and cannot be recovered if the firm exits the industry. In the second stage of the game the firms choose prices i.e. they play a "Bertrand" game. The timing of such a game reflects the idea that the entry decision is a long-run decision whereas price decisions are short-run ones. The equilibrium concept for such a multi-stage game is the standard sub-game perfect Nash equilibrium and is found by first looking at the equilibrium of the last stage and then moving backwards (see chapter 8). I focus on pure strategies.

To find the solution start from the price sub-game. Here there are three possible cases. First, both firms have entered in the previous stage. Because of price competition, they will earn zero profits. Second, one firm only has entered. Hence this firm earns the monopoly profit Π_M (which we assume to be larger than the fixed cost f) and the other firm zero profit. Third, no firm has entered the industry. Profits are zero for both. Once found the solutions for the last stage of the game we can now pass to the first stage.

The payoff matrix at the first stage of the game is summarised by Table 2.1.

INSERT Table 2.1 The entry game

Note that when both firms enter they incur a loss due to the fixed set-up cost. There exist two equilibria in pure strategies of the game. One (E,NE) where firm 1 enters and firm 2 does not, and the other (NE,E) where firm 2 enters while firm 1 does not. Consider first the pair (E,NE). For this to be an equilibrium, each firm must not have an incentive to deviate from it for given strategy of the other player. If firm 1 deviates from the strategy "enter" given that firm 2 does not enter, it would attain a payoff $0 < \Pi_M - f$. The deviation is clearly unprofitable. The other possible deviation is that firm 2 chooses the strategy "enter" given firm 1's decision to enter. In this case, however, firm 2 would receive a payment $-f < 0$. It would be worse off with respect to the zero profit associated with the decision of not entering. This deviation is not profitable either. Therefore, the pair (E,NE) is an equilibrium. Since the game is perfectly symmetric, one finds that the pair (NE,E) is also an equilibrium.

No matter which firm ends up with being the only producer, a monopolistic structure arises in the market. Despite free entry, the market outcome is a monopoly, and the monopolist does earn supra-competitive profits. This simple example shows that the anticipation of market competition might lead one firm not to enter the market even when entry is not restricted. The same argument is developed in what follows, in the context of a more sophisticated setting.

Finiteness property in vertical differentiation models* One of the most important recent contributions in oligopoly models is given by the work of Shaked and Sutton (1982, 1983, 1987). They analyse a *vertical product differentiation* model, that is a model where consumers agree on the ranking they give to products of different qualities. Whereas everybody agrees on the quality of the goods, differences in incomes exist which do not necessarily allow each consumer to buy the best products. Shaked and Sutton (1982, 1983) analyse a game where firms choose whether to enter the industry or not at the first stage, choose quality of the good at the second stage, and choose price at the third stage. They show that the number of firms which can co-exist in an industry described by the aforementioned characteristics is bounded above even when the size of the market tends to infinity. In the limit, if the distribution of incomes is narrow enough, only one firm will operate in the industry despite free entry and a number of potential firms. Also, the monopolist will earn supra-normal profits. This result is in sharp contrast with the standard homogenous good with Cournot competition and the *horizontal product differentiation* models (that is, models where consumers differ in their preferences over distinct varieties of the same product), where an increase in market size allows more firms to cover their fixed costs, leading to a more fragmented market structure.

The intuition for the finiteness property result, given in very loose terms, is as follows. Imagine that a new firm considers entry in an industry monopolised by a firm producing a good of quality u_1. If the new entrant chooses a quality $u_2 < u_1$, then it will have to offer the good at a lower price as well, since consumers have a preference for quality. However, competition would bring down the price of the high quality good, making it more attractive to all the consumers. In the extreme case where people have similar enough incomes (or, equivalently, tastes for quality), all of them would continue to buy the top quality good, despite the appearance of a new quality. Note also that given that there exists price competition in the market, firms would never choose a similar quality to the existing one, since this would increase competition in the marketplace. Producing identical qualities would make the goods homogeneous, and the game at the last stage would be like the Bertrand one, with zero profit at equilibrium. The new entrant might be able instead to enter with a higher quality than u_1, but in this case it would be the incumbent firm which would have to leave the market. From the point of view of the final result on welfare, it would not make any difference which one is the firm that stays in the market at equilibrium. However, it is unlikely that the entrant would be able to replace an incumbent, since the latter has a strategic first-mover advantage.[62]

The finiteness property result rests on many assumptions, of which the strongest is possibly the following: the costs of producing a higher quality fall on fixed costs such as R&D investments and advertising, rather than on variable costs such as better raw materials and more expensive high-skilled labour. If this assumption is relaxed, then the top quality firm will have higher variable production costs than a low quality firm, and this restricts its possibility to decrease prices. Under fixed costs of quality and identical marginal costs, a high quality firm has a better ability to set prices which are as low as those set by a low quality firm and steal the market share of the latter (recall that at identical prices everybody would buy the high quality product). But under variable costs of quality, the top firm has higher marginal costs as well, and this reduces its ability to compete on prices. Therefore, the finiteness property does not hold when variable costs of quality are relatively more important than fixed costs of quality, as showed in Shaked and Sutton (1987).

Shaked and Sutton (1987) and Sutton (1991) reformulate the finiteness property result in a weaker but more robust version, in the sense that it

[62] See Motta, Thisse and Cabrales (1997) for an analysis of a quality race between firms which start from different levels of quality. This paper shows that persistence of leadership is the likeliest outcome in this type of games.

holds across a number of specifications, and in particular when firms produce goods which are not only qualitatively different, but are also horizontally differentiated. According to this version, as the size of the market increases, a very large number of firms may survive in the industry, but there always exists at least one firm whose market share is significantly bounded away from zero (in the homogenous good case instead, when the number of firms tends to infinity, their market share tends to zero).

Sutton (1991) puts the model to a test, and gathers impressive econometric, as well as case-study, evidence in favour of this result. In particular, the key role played by advertising expenditures (which allow the firms to increase the perceived quality of the product for the consumers) emerges clearly. In most of the sectors analysed (food and beverage industry), firms have been able to conquer large market shares by escalating advertising expenditures. When market size increases, advertising outlays of the firms tend to increase as well, since there exists an incentive to increase quality (the marginal cost of quality-enhancing spending is unchanged, but the marginal benefit is higher because of the larger market size). This increases the fixed sunk costs of the investments needed, and tends to limit the number of firms in the market. More precisely, new entrants might still enter the industry and manage to operate profitably, but they would have to rely on a low-price, low-quality strategy, which does not jeopardise the market positions of the incumbents.[63]

6.3.2 Switching costs

Another situation where market power does not necessarily decrease under free entry is when there exist consumer switching costs.[64] There are many reasons why consumers might prefer to stick to products already bought in the past, other things being equal. Switching to a new product (or a new supplier) might entail transaction costs (when one closes an account in a bank and opens another in a new bank, for instance) and learning costs (the cost of passing to a new software application, after having learned how to operate

[63]It should be emphasised that advertising expenditures are not an exogenous sunk cost but rather an endogenous variable for the firm. If they were fixed sunk costs, namely exogenous barriers to entry, then market size increases would allow a larger number of firms in the industry and a more fragmented market structure. However, this occurrence is rejected by the data.

[64]For models with switching costs, see Klemperer (e.g., 1987a and 1987b). For surveys on this topic, see Klemperer (1995) and Padilla (1991).

with a different one, for instance). Some of these switching costs might also be artificial or contractual, that is created on purpose by the firms in order to make it more difficult for consumers to pass to new products. Think of "frequent-flyer" programmes, for instance, that make travellers keener on continuing to use the services of the same airline: if after having accumulated miles with an airline, they switched to another, they would not reach the number of miles necessary to have a free flight. Another example of artificial switching cost is fees charged by banks to close an account.

In all these cases, the existence of switching costs effectively differentiates goods which would otherwise be perceived as perfectly identical. One might be perfectly indifferent, *before* opening a bank account, between two banks that charge similar rates and give similar services. However, *after* having opened an account at a particular bank, the existence of switching costs would make it worth changing bank only if the alternative bank will give much better rates or services. Products that are ex-ante identical after a purchase become ex-post differentiated.

When such switching costs exist, and one can realistically think that this is the case for many industries, new entrants generally have a harder time in getting market shares from the incumbents. Firms which have already developed a large base of customers will have a large advantage, since very important price cuts should be offered by new firms to attract committed customers. Again, free entry does not guarantee that market power will decrease.[65]

Switching costs might also allow incumbent firms to choose pre-entry prices and quantities so as to strategically deter entry in later periods. For instance, a firm that is a monopolist today might under-price in order to build a large customer base and make it more difficult for a potential entrant to enter the market tomorrow.[66]

[65] However, switching costs might sometimes make the incumbent firm less keen on fighting entry and more tempted instead to set high prices to exploit its customer base. In other words, prices will be higher relative to a situation where there are no switching costs, and this in turn makes entry more likely other things being equal.

[66] In some circumstances, for instance when tomorrow's demand is much larger than today's, a firm might instead want to overprice today, thereby committing to more aggressive behaviour tomorrow. See Klemperer (1987c) for an analysis of entry deterrence behaviour.

The competitiveness of switching cost markets Switching costs are not necessarily always as anti-competitive as they might appear at first sight. Suppose for instance that there are two firms which are both new entrants in an industry where consumers will have some kind of switching costs. In each period, each firm has simultaneously to decide the price to be charged in the market. Then the degree of competitiveness in each period will be very different. In the second period, each firm will have some captive customers: the consumers who have already bought from the firm in the first period will tend to buy from the same firm in the second period too (they are locked-in, at least partially). Therefore, we should expect their lower elasticity of demand with respect to price to be reflected in higher prices charged by the firms in the second period. Notice that the larger the first period market share of a firm the larger the profit it will make in the second period.

In the first period, however, we should expect stronger competition than in a situation without switching costs. Since the first period market share (or customer base) has a positive impact on the second period profit, each firm will price more aggressively in the first period. Overall, therefore, it is not possible to conclude whether switching costs will result in higher or lower total welfare. In two-period models, the net effect on competitiveness due to switching costs is therefore ambiguous.

However, the presumption that switching costs decrease competition is reinforced by the analysis of models where firms interact not for only two but for many periods. Beggs and Klemperer (1992) examine a market where, in each period, firms have to set prices, some new consumers arrive, some old consumers leave the market. They show - under relatively general assumptions - that in such a situation switching costs relax competition, that is, prices are higher than in markets without switching costs. Intuitively, there are two different effects at work when firms set prices in any current period. On the one hand, they would like to charge higher prices to exploit the current customer base. On the other hand, they would like to set lower prices to expand the future customer base (in order to be able to exploit it later). For several reasons one should expect the former effect to dominate. First, since today is more important than tomorrow, it is better to have higher profit today than tomorrow. Second, lower prices today are relatively ineffective in attracting consumers, since consumers would correctly anticipate that they will face higher price tomorrow. Hence, it is unlikely that today's prices will be much lower to attract future customers.[67]

Conclusions Although an unqualified proposition is not possible, especially in the light of the various strategic effects that they bring into play, switching costs are generally detrimental to welfare, among other things because they make entry more difficult and markets less competitive. After a long and detailed review of switching costs and their effects, Klemperer (1995) arrives at the conclusion that:

> "public policy should discourage activities that increase consumer switching costs (such as airlines' frequent-flyer programmes), and encourage activities that reduce them (such as standardization that enhances compatibility and reduces learning costs of switching, and quality regulation and information sources that reduce consumer uncertainty about untested brands)." (Klemperer, 1995: 536)

It would be difficult to argue that competition authorities should systematically forbid firms from introducing contracts and practices which involve switching costs, among other reasons because some of these practices might conceivably have reasons other than the desire to reduce market competition. However, authorities should check that firm-created switching costs are not preventing competition in markets. When de-regulating a previously monopolised sector, for instance, authorities should make sure that consumers are not locked-in by artificial switching costs (like when switching telephone provider would imply changing one's telephone number). When analysing mergers, authorities should use their bargaining power to impose reduction on switching costs which are artificial and not motivated by any possible transaction cost savings.[68]

[67] Additionally, one should expect that in most situations prices will be "strategic complements" across periods. If a firm charges a higher price today, its rival will have a larger customer base today. Therefore, it will be more likely to charge higher prices tomorrow, since it will want to exploit its larger customer base. In turn, higher rival prices will make one firm's profit larger.

[68] In mergers of airlines, for instance, the European Commission is right in asking the merging companies to make smaller rival airlines also benefit of their frequent-flyer programmes. See for instance the *SAS/Lufthansa* and *Swissair-Sabena* cases.

6.3.3 Network effects

Other industries where a monopoly might persist despite the absence of barriers to entry are those characterised by network effects. In such industries, consumers derive utility from the number of other consumers who choose the same product. If most consumers have already bought a given product, it will be difficult for new firms to attract demand.

Network effects are mainly of two types.[69] The first type is given by *physical*, or communications, networks. In this case, a consumer's utility in the consumption of a good increases *directly* with the number of other people consuming the same good. The prototypical example for such a network is telephones. One would not do much with a telephone if there were no other people with telephones with whom one could communicate. The larger the number of people in a given telephone network the more useful the telephone service. Fixed telephones, mobile telephones, fax, telefax, telex, e-mail services clearly belong to the same category.[70] The second type is given by *virtual*, or hardware-software, networks. In this case, a consumer utility increases *indirectly* with the number of other consumers buying the same good because of its effects on the availability of a complementary product. Think for instance of a credit card network. As a card-holder, my utility is not directly affected by the number of other consumers using the same type of credit card. However, the larger the number of holders of the same credit card the more likely that shop-keepers will accept it. Similarly, a consumer's utility from a given computer hardware (or any type of durable good, be it a car, a washing machine, or a VCR) increases with the number of other buyers of the same computer (or durable good), since this will increase the likelihood that computer software for it will be developed (or spare parts, post-sales services, or video cassettes will be available).

In all these situations, consumers will have to face coordination problems, since their choices are based on what other people will also do. In some cases, this might not be a problem. For instance, nobody will wonder today whether there will be enough people with whom to communicate by

[69] The first articles on network effects were Katz and Shapiro (1985), and Farrell and Saloner (1985). For a survey, see for instance Katz and Shapiro (1994), and for a simple introduction to these topics, Shy (2001).

[70] But communication might not be necessary for a direct increase in utility. Think for instance of some consumer products, where the perceived value of wearing a certain type (or brand) of shoes or jacket or watch, often depends on many other people also having the same product.

fixed-line phones or e-mail. But in other cases, when completely new networks/products are introduced, expectations about what other people will do are relevant. The development of a new generation of mobile phones might be hindered by doubts about how many other consumers will adopt them (and how quickly). Likewise, when a new firm launches a network with a different standard than the one currently used, a potential buyer has to guess the likelihood that such network will be successful, in order to avoid being locked-in with a product that gives her little possibility of communication (communications networks) or little software/services/complementary goods (hardware-software networks).

Therefore, expectations play in network industries a crucial role. Suppose for instance that a new type of mobile phone services were introduced by an entrant firm, whose standard is incompatible with existing mobile phones. Each prospective customer has to take a purchase decision on the basis of her expectations of other prospective customers. If all consumers have reasons to believe that million of other people will rush to buy the new type of phones, all will buy. If none thinks that anybody would buy it, none will. Notice that here expectations will be self-fulfilling: in the first case, the service will be successfully established; in the second, it will not.[71]

These features of network effects explain why a potential entrant might find it very difficult to challenge an incumbent in these industries. It is not enough to have a better product, or to provide it at a lower price, since a crucial component of the utility of people is given by the number of (current and future) users of it. If the new product is not compatible with the established one, the firm has to convince prospective buyers that enough other buyers will buy it. The larger the number of consumers already locked-in with the current standard the more difficult will be its task. The stronger the reputation of the new entrant and the more resources it commits to the new product the higher chances that it will succeed. A number of strategies might be used to such purposes, from introductory price offers (or even giving away the product for free), to convincing firms selling complementary

[71] Satellite-based mobile phones failed to attract enough consumers, and the firm providing its services went bankrupt. GSM services in the US have developed very slowly (relative to Europe), among other reasons because of uncertainty on the winning standard. DIVX, a technology offering similar services as DVD, but based on a different standard, also failed in attracting enough customers and disappeared from the market. On the last case, see Dranove and Gandal (2000) for a description on how the uncertainty of consumers about the dominant standard can affect the evolution of the market.

services (or spare parts, or software applications) to develop them.

However, incumbents can also adopt a number of strategies which might delay or completely deter new entrants. First of all, incumbents will want to make sure that the new products cannot be compatible with theirs (on compatibility and inter-operability, see also chapter 7). As long as a standard is proprietary, this strategy will be legal. However, the incumbent might also engage in anti-competitive practices. For instance, faced with an entrant which offers a product with new features, an incumbent might want to announce that soon it will introduce an upgrade of its product that incorporates these new features even if this is not true.[72] Likewise, it might advertise that the entrant is making very slow progress in building a customer base (or equivalently, that the incumbent's market share is high and growing). Such announcements will likely have an impact on the expectation of consumers about the viability of the entrant, and should therefore be carefully monitored by the anti-trust authorities, and promptly punished if false.[73]

Given that entry might be difficult under incompatible standards, another route that antitrust authorities might conceive to avoid such problems is to force compatibility in the industry. For instance, they might impose on an incumbent full inter-operability with an entrant's products. However, this solution reminds us of the discussion on patents and essential facilities (see above). Ex-post (that is, after a given standard is dominant), the imposition of inter-operability sounds beneficial because it allows more competition. However, ex-ante (that is, before a given product appears on the market) it has an adverse effect of innovations, since it discourages firms from introducing new products and to fight for them to become the standard in the industry. Therefore, such measures do not appear convincing in general.[74]

A full analysis of the welfare effect of industries with network effects is beyond the scope of this section. Nevertheless, it is important to deal

[72] Product pre-announcements can strongly affect consumers' expectations, but it might be difficult to check to what extent they are false, since they refer after all to intentions and plans...

[73] For a discussion of anti-competitive practices in network industries, see also Rubinfeld (1998).

[74] For a more thorough discussion of public policies about compatibility, and their possible effects on innovations, see Farrell and Katz (1998). An issue they also discuss is that interfaces that do not involve much innovation efforts should not receive protection. For instance, imposing *number portability* to telephone companies rightly ensures that the interface "telephone number" is not protected (assigning telephone numbers to consumers is certainly not worth protection).

- albeit briefly - with the idea that such industries are particularly prone to dominance, and are therefore naturally associated with the existence of monopolies. The basic concept here is that of *market tipping*, which refers to the fact that when there are competing systems, once a system manages to gain a certain advantage in consumer preferences, then it might become more and more popular (think of the role of expectations discussed above: once consumers see that a product gains a large enough market share, then it is expected to become the industry's standard) and its rivals might fade out.

Market tipping is certainly an important phenomenon in network industries, but some qualifications should be made before jumping to general conclusions. First, there are many situations in which different standards might co-exist in a given industry. For examples of incompatible systems, think of different credit card and ATM networks,[75] or different brands of video games. In many cases, consumers do value variety and differentiated systems will survive in the same industry. Second, the existence of tipping and of the large profit that can be reaped once the own product is established as the industry's standard, will prompt firms to compete fiercely to win the "standards war". Intense promotional activities of various types, as well as aggressive pricing in the introductory periods, might characterise the initial stages of a given product life, as a firm attempts to increase market shares so as to gain the edge it needs to make the market tip.[76] As a result, it might well be that the large profit made by a firm after its product has become the dominant standard might just cover the cost incurred during the standards war.[77] Third, the existence of only one network in the market might even benefit consumers, to the extent that they will be able to enjoy more communications possibilities or more complementary services, whereas under competing networks they will not be able to.

Likewise, the difficulties faced by entrants in markets which exhibit important network effects should not be generalised to think that all these

[75] In some countries, there exist different networks of Automatic Teller Machines.

[76] Rey, Seabright and Tirole (2001) stress a "topsy-turvy" principle of competition. The higher the profits to be made in a market the stronger the competition to obtain them. This should hold in industries with network effects, switching cost, lock-in effects, economies of scale, and generally all situations where the "winner takes all".

[77] This reminds us of the analysis of the competitiveness of markets with consumer switching costs. Indeed, switching costs and network industries share a number of similar features, as stressed in Farrell and Klemperer (2001).

markets will naturally show "excess inertia" (or persistence of dominance). As Katz and Shapiro (1994, p.108) note, examples of entrants which manage to establish a product incompatible with previous standards abound in the real world. Further, from the welfare point of view, it is not even clear that entry - if occurred - would always be beneficial. Suppose for instance that - due to consumers' expectations, perhaps helped by the reputation of the entrant - a new standard which does not offer advantages over the previous one affirms itself in the industry. This will lead to "stranding" of the buyers who have locked themselves in the old technology: their purchases will rapidly become obsolete, and duplication of purchases will have to follow.

6.3.4 A model of (physical) networks*

In this section I present a simple model of physical (communications) networks that is proposed by Fumagalli, Karlinger and Motta (2003), to illustrate the coordination problems that arise for consumers when network externalities exist, and show that this might result in the entrant staying out of the industry even when it is more a more efficient producer.[78]

Following Katz and Shapiro (1985), assume that consumers value a network good i as follows:

$$U_i = r + v_i(n) - p_i, \tag{22}$$

where r is the intrinsic value of the good in the absence of network effects, $v_i(n)$ is the valuation for the network n consumers buy network i good), and p_i is the price a consumer has to pay to join network i. For simplicity, assume that $r = 0$, so that the network good has value only for its "network" component.

Assume that $v_i(n)$ is non-decreasing and concave, but that after a certain number of consumers z the network exhausts its positive externalities. In other words, $v_i(z) = v_i(z + j)$ for all $j > 0$. (The reason behind this assumption will be explained shortly.) Assume also that $v_i(1) = 0$: there is no value from buying a network product of which one is the only consumer.

In the market for this network good, there is an incumbent, I, and a potential entrant, E. The entrant is more efficient and can produce at a

[78] Fumagalli et al. (2003) develop the model to analyse how the incumbent and the entrant might resort to different strategic variables so as to make consumers coordinate on (resepctively) an exclusionary or an entry equilibrium.

cost $c_E < c_I$. The two networks are homogenous in the sense that if they are of the same size, they create the same externality for their consumers: $v_I(\cdot) = v_E(\cdot) = v(\cdot)$. Assume also that the entrant has arbitrarily small fixed cost of entry, ε, to underline that entry barriers come only from network effects.[79]

There are two groups of consumers, each of size z. The first group of ("old") consumers have already bought the incumbent's product in the past, and enjoy a utility $v(z)$. The second group of ("new") consumers is on the market now.

Note that the assumption that the incumbent has already served z consumers in the past has a two-fold role. First, it implies that new consumers, should they join the incumbent's network, do not have any externality on the z old consumers; second, it implies that the asymmetry between the incumbent and the entrant is maximal, since the incumbent has already reached the highest possible externality level.

The game is as follows.

1. E decides whether to enter or not.

2. Active firms set prices simultaneously. Assume for the time being that prices must be the same for all customers.

3. Buyers decide which network to join, and pay either p_I or p_E.

Two types of equilibria There exist two very different equilibria in this game.[80] The first one is an "entry" equilibrium, where the entrant enters and all new consumers join its network. The second one is an inefficient "persistence of monopoly" equilibrium, where consumers fail to coordinate on the outcome that would be more efficient. Let me characterise and prove the existence of each equilibrium as follows.

Entry equilibrium. At this equilibrium, E enters, I sets a price c_I and E a price a shade below c_I, and all z new consumers join the entrant's network.

To see why this is an equilibrium, consider consumers first. At this equilibrium, they have a surplus $v(z) - c_I$. If one of them deviated and bought

[79] Note that here networks are incompatible: there exist network externalities only among consumers who buy from the same network. See chapter 7 for a model where inter-operability (or compatibility) is an endogenous choice of firms.

[80] There might also exist other equilibria.

I's product, when all others are joining E's network, it would get a utility $v(z) - c_I$. Therefore, it has no incentive to deviate.

Firms have no incentives to deviate either. Firm I would have no incentive to decrease its price given that, should it get consumers, it would make losses. Firm E would not have an incentive to increase its price because otherwise customers would switch to the incumbent.

Persistence of monopoly (or mis-coordination) equilibrium. At this equilibrium, E does not enter, I sets a price $p_I = v(z)$, thereby extracting all surplus from consumers, and all z new consumers join I's network.

To see why this is an equilibrium of the game, consider the case where the entrant has entered and has offered a price which might be as low as c_E. It is easy to see that everybody addressing the incumbent is still a Nash equilibrium at the consumers' stage of the game. At the proposed equilibrium, all consumers buy the incumbent's product, and have surplus 0. However, if any of them deviated and bought from E, given that all others buy from I, it would receive utility $v(1) - c_E = -c_E$. Therefore, it would worsen its position.

As for the firms, at this equilibrium I has no incentive to change its price (it is earning monopoly profits), and E has no incentive to deviate, as by entering it would not be able to recover its fixed cost, however small.

This is clearly a mis-coordination equilibrium, in the sense that consumers are unable to coordinate on the outcome that is more convenient for them. Were they able to coordinate and decide together, the only equilibrium of the game would be the one where entry occurs. (I shall come back to mis-coordination problems in chapter 5, when discussing buyers' power, and in chapter 6, when discussing exclusive dealing as an entry deterrence device.) On the other hand, the incumbent might also to resort to strategies that allow it to exclude entry, as I now show.

Price discrimination to exclude entry in a network industry Assume now that the incumbent can price discriminate among the new consumers, and that its costs are not too low, i.e. $2c_I < v(2)+c_E$. For simplicity, set $z = 2$, and call the new consumers 1 and 2. I now show that even if the entrant entered and offered its network product at a price c_E, it would not get any demand if the incumbent is able to price discriminate.[81] Therefore,

[81] The incumbent would be able to deter entry also if it consumers bought sequentially. See also chapter 6 for an exclusive dealing model which exhibits very similar features.

the only equilibrium of the game is the one with persistence of monopoly.

If the entrant sets a price c_E, and the incumbent sets prices $p_1 < v(2)$ and $p_2 = v(2)$, consumers' payoffs will be as in Table 2.2.

INSERT Table 2.2: Consumers 1 and 2's payoffs under price discrimination.

The equilibrium where both consumers buy from the incumbent always exists: consumer 1 has a surplus buying from I, whereas it would have a negative surplus if it deviated and bought from E (recall that a network of size 1 gives zero utility); consumer 2 has zero utility anyhow.

However, the entry equilibrium, which corresponds to the pair (*buy* E, *buy* E) does not exist any longer. To see why, note that, given that 2 buys from E, consumer 1 gets surplus $v(2) - p_1$ if it buys from I, and $v(2) - c_E$ if it buys from E. Therefore, it will want to deviate and buy from I if $p_1 < c_E$.

But is it profitable for the incumbent to deter entry? Clearly, it would not be able to profitably deter entry if it had to offer a price below cost to each buyer. However, under price discrimination it offers a price below cost to one buyer only, while setting the monopoly price to the other. Therefore, this entry-deterrence strategy is profitable as long as total revenues, $v(2) + p_1 \geq 2c_I$ are higher than total costs. Since $p_1 < c_E$, this profitability condition becomes: $v(2) + c_E \geq 2c_I$, which is the assumption made above.

6.3.5 Exclusionary practices

As we have seen, free entry does not guarantee that an industry structure will become less concentrated over time. In markets characterised by endogenous sunk costs, consumer switching costs or network effects, for instance, entrants will often find it difficult to challenge successfully incumbents, even if the latter do not behave strategically.

When incumbents behave strategically, things turn even more difficult for entrants. A monopolist (or more generally a firm with large market power) might engage in many different practices aimed at deterring entrants. Investing in extra-capacity, setting prices below cost, flooding a market with many different product specifications, foreclosing access of rivals to crucial inputs, bundling, price discriminating, tying are all possible examples of strategies that can prevent entry. The analysis of exclusionary practices is delayed until chapter 7, but it is worth recalling here that competition authorities should be vigilant and promptly intervene whenever monopolists impede entry through practices whose profitability derives only from their ability to keep entrants off the market. This is an important issue, as well as a difficult one, since in most cases it is hard to tell apart genuine competitive strategies from predatory ones.

7 Summary, and policy conclusions

This chapter has illustrated the relationship between market power and welfare. The analysis of allocative efficiency has showed that market power brings about a welfare loss, due to higher prices than in a competitive situation. Productive and dynamic inefficiencies (higher production costs and lower innovation rates) might also be associated with market power. This explains why competition policy should be concerned with market power.

However, I have also argued that the elimination of market power - even if it were practicable - is not one of the objectives competition policy agencies should pursue. Indeed, the prospect of having some market power (i.e., some profit) represents the most powerful incentive for firms to innovate and invest. Competition laws and their enforcement should therefore ensure firms will be able to enjoy the rewards for their investments. I have therefore argued that any expropriation of firms' assets (whether material or immaterial) should be avoided. As a consequence, resorting to the doctrine of essential facilities (granting access of crucial assets to competitors), to price controls, or even more drastic structural remedies must be done only in truly exceptional circumstances.

I have also tackled other misconceptions of competition policy. In particular, I have underlined that defending competition is not tantamount to defending competitors. Indeed, competition often leads inefficient firms to exit and this is beneficial from a welfare point of view. Protecting inefficient firms so as to prolong artificially their life in an industry would be detrimental.

Finally, market forces left alone will not "fix it all": for several reasons, very often incumbent firms are able to keep and reinforce their market power. Competition policy has therefore the important role to be vigilant, and guarantee an environment where potential and actual competitors are able to challenge the firms enjoying a position of large market power.

8 Essays and exercises

Exercise 1 *[82] *An example of rent-seeking. Consider a market where a total profit Π can be earned by the firm which obtains the monopoly right to sell in that market. Suppose that there exist n identical firms which participate in the competition to obtain this monopoly right. Each firm i has to simultaneously decide the amount x_i it wants to spend, knowing that the probability to win the race is given by $x_i / \sum_{j=1}^{n} x_j$. (a) Find the symmetric equilibrium level of expenditure of each firm, and the expected equilibrium profit of each firm. (b) Show that as n tends to infinity, the total expenditures made by the firms equal the total monopoly profit.*

Exercise 2 *By making use of graphs, identify the welfare loss occurring in a market where the price p is higher than the marginal cost of production, and then discuss how this loss could be higher due to possible rent-seeking activities and productive inefficiency.*

Exercise 3 *Explain why there is a possible tension between allocative efficiency on one side and productive and dynamic efficiency on the other side.*

Exercise 4 **Inspired to d'Aspremont and Motta (1990). Consider a homogenous good industry with two potential firms. Market demand is given by $p = S(1-p)$, where S is market size and Q is industry output. Firms have zero constant marginal costs but incur a fixed cost $k \in (0, S/9)$ if they are active. The game they play is as follows. First, they decide whether to enter or not the industry. Then, they compete in the product market. Consider three different cases of product competition: (i) firms non-cooperatively choose prices (Bertrand); (ii) firms non-cooperatively choose quantities (Cournot); (iii) a cartel is allowed: firms set quantities (or prices, it is equivalent) so as to jointly maximise their profits (Monopoly). Focus on pure strategies and find (sub-game perfect Nash) equilibrium quantities, prices, profits, consumer surplus and welfare for each of the three cases above. Show that for low enough fixed costs Cournot competition gives rise to higher welfare than Bertrand competition, and that Monopoly always gives the lowest welfare level at equilibrium.*

Exercise 5 **Consider the same example as in section 4.3.1, but assume that there is a drastic innovation, that is the reduction in marginal cost is large enough: $x > 1 - c^h$. Compare the equilibrium under duopoly with the one obtained above under monopoly.*

[82] I am grateful to Joe Harrington for suggesting this formalisation.

Exercise 6 ***Consider a variant of the model dealt with in section 4.3.1 where the amount of innovation is continuous and endogenously determined rather than exogenously given. The cost of innovation is not constant, but is a quadratic function of the amount of innovation made. More explicitly, a firm faces an investment cost $C(x_i) = x_i^2$. By investing x^2 in innovation (such an investment still occurs in the first stage and it is therefore sunk in the second stage), a firm will have marginal cost $c^h - x$ in the following price sub-game. Apart from this variant, consider the same set of assumptions as above. (1) Find the optimal amount of investment x_m for a monopolist. (2) Find the equilibrium amount of investment x_{nd} in the duopoly. Does it corresponds to a drastic or non-drastic innovation? Is x_{nd} bigger or smaller than x_m?*

Exercise 7 *Why should one expect an inverted U relationship between competition and innovation?*

Exercise 8 *** Consider a homogenous good industry with perfectly symmetric firms. The firms play the following game. In the first stage, they simultaneously decide whether they enter the industry or not. In the second, they simultaneously decide on the quantity to bring to the market. If they enter, they have to incur a fixed set-up cost F. Production occurs at a constant marginal cost c. Market demand is characterised by $p = 1 - Q$, p being the market price and Q the aggregate output. (a) Find the number of firms n that would enter at the sub-game perfect Nash equilibrium. (For simplicity, consider n as continuous in the whole exercise.) (b) Consider now the following game. In the first stage, a social planner decides the number of firms (and pays the fixed costs of firms). The social planner maximises welfare, as given by the sum of consumer surplus and producer surplus minus the entry costs it has to pay. In the second stage, the same quantity competition game as above is played. Find the optimal number of firms chosen by the social planner and compare it with the number of firms that would enter at the sub-game perfect Nash equilibrium.*

Exercise 9 *** Consider the model analysed in section 4.3.2. Assume that there exists a social planner who maximises total welfare taking the number of firms n as given. First the social planner has to choose R&D levels, and then output. Find the (symmetric) optimal levels of R&D and output, and compare them with the equilibrium levels obtained in the model of section 4.3.2.*

Exercise 10 ** Consider the same model as in section 4.3.2 but assume that the n firms jointly determine their output levels: in the first stage of the game, each firm independently sets its R&D level; in the second stage, the firms act as a cartel, that is they choose their output levels so as to maximise joint profits. Find the equilibrium levels of R&D and output and compare them with those obtained in the model of section 4.3.2.*

Exercise 11 *Industrial organisation models suggest that - left alone - market forces will result in any given industry in equilibrium levels of quality, variety, innovation, or competitors that are different from the optimal levels for society (that is, from the levels of quality, variety, innovation or number of firms that would maximise economic welfare). This is because firms take their decisions on the basis of their profits, not taking into account the externality that such decisions exercise over competitors, or over consumers. What are the possible practical implications of this result? Can competition authorities try to move each industry towards the optimal outcome?*

Exercise 12 *In a country, the fixed telephone market has just been deregulated and new entrants are allowed to challenge the incumbent monopolist telecom operator. Soon, the entrants discover that it is very difficult to persuade customers to switch operator, because they do not want to change their telephone number, over which the incumbent claims intellectual property rights. Should the Court decide in favour of the incumbent (in which case, customers who patronise the new operators should change telephone number) or the entrants?*

Exercise 13 *(Loosely inspired to the Oskar Bronner case) The leading national newspaper in the country has established a system of home delivery services, that allows it to distribute copies of the newspaper in the early morning in all the country. A regional newspaper has decided to sell over the all national territory, and accordingly increases coverage of news, from local to national. However, it discovers that the lack of a home delivery distribution makes it difficult to increase national sales, and that building such a distribution system is too expensive and risky. It asks the leading newspaper to grant it access to its home delivery distribution system behind a fair compensation, but the latter refuses it. Should the Court decide in favour of the national newspaper or the local one?*

Exercise 14 ***Leasing v. selling (From Tirole, 1988, section 1.5.2.1) Consider a two- period game where there exists a monopolistic firm that wants to sell its durable good. The durable good will last only two periods, and after that it will become obsolete. There is no depreciation of the good between the two periods. The discount factor δ is identical for all consumers and the firm. Demand for the utilisation of the good will be given by $p = 1 - Q$ (Q being aggregate quantity). Production is assumed to be costless. A resale market exists: consumers who buy the good in one period might want to re-sell (or lease) it in the second period.*

1. Consider first the case where the firm sells in each period. Find the equilibrium prices charged by the monopolist in each period, and show that they decline over time. Find the total equilibrium profit.

2. Consider then the case where the monopolist leases (rents) the good in each period, and find equilibrium prices and profits. Are they higher or lower than the profit made under sales?

Exercise 15 *Vertical product differentiation models suggest that high concentration might arise in markets characterised by endogenous sunk costs such as, for instance, advertising outlays. In such markets, firms investing heavily in advertising typically have larger market shares, charge higher prices and get higher profits than their rivals. Their leadership also tends to persist over time. What policy implications should we derive from such models?*

Exercise 16 *** (From Klemperer, 1988: Welfare effects of entry) Consider the following two-stage game between an incumbent firm and a potential entrant. In the first stage, the entrant decides whether to enter or not in the market. Entry is costless. In the second stage, existing firms in the market (simultaneously) choose the quantity to be sold. Firms have the same constant unit cost c. Consumers who buy the entrant's product will have to incur a switching cost s. (For simplicity, assume that there is no switching cost when buying from the incumbent.) Assume that there is linear demand $p = \alpha - \beta q$ (in other words, this is the q-th consumer reservation price net of switching costs).*

1. Find the equilibrium output and prices for the game according to the different levels of switching costs s.

2. Is it possible that entry diminishes welfare?

3. Consider the case where the switching cost is zero. Would welfare be higher or lower than under positive switching costs?

Exercise 17 ****(Examples 0 and 1 in Klemperer, 1995.) Consider the following two-stage game. Assume that there are N consumers who are distributed along a line $[0,1]$ which measures their (linear) cost of learning how to use a (homogenous) product. Firms A, B sell the same product but are located at 0 and 1 respectively, and have the same unit cost $c = 0$ in each period. Therefore, a consumer located at point y has a learning cost ty of using firm A's product and $t(1-y)$ of using B's product ($t > 0$). Consumers do not have any physical transport cost. At period 1, consumer utility is given by $U = r - p_i - t\,|l_i - y|$, for $i = A, B$, where l_i stands for location of firm i. Assume also that goods cannot be stored and that consumers have the same discount factor $\delta = 1$. At period 2, each consumer has a reservation price R to buy the good. Goods are perceived as perfectly homogenous, but there are switching costs: if one changes provider, one has to pay a cost s which for simplicity is assumed to be independent of the distance which separates the consumer from the firm. Firms set simultaneously prices in each period. Assume also: $R > c$; $s \geq R - c$; $r - 2t > c$.*

1. Find the equilibrium prices of each period of the game.

2. Find the equilibrium price for the game where the first period is like above, but in the second period there are no switching cost at all. Show that under switching costs prices are higher in the second period but lower in the first period.

Exercise 18 **Monopoly threatened by entry. (From Tirole, 1988: section 10.1.4) Consider the case of two firms in an output market. Consider also that there is a third firm which cannot produce in this output market but has generated an innovation which can lower the unit production cost from $\overline{c}$ to $\underline{c} < \overline{c}$ (process innovation). This third firm puts this innovation up for bidding between the two producing firms and the innovation is protected by a patent of unlimited duration. Before the bidding process takes place, firm 1 is a monopolist and produces at unit cost $\overline{c}$. Firm 1's profit at this status quo market structure is denoted by $\Pi^m(\overline{c})$. Firm 2, is a potential entrant whose unit cost is infinite. Let $\Pi^d(\overline{c}, \underline{c})$ and $\Pi^d(\underline{c}, \overline{c})$ denote the profits for the monopolist and the entrant, respectively, if the entrant alone adopts the new technology with marginal cost $\underline{c}$ and, consequently, the monopolist's marginal cost is still $\overline{c}$.*

1. Find the value of the innovation (i.e., the difference between the profits a firm earns if it is the one that wins the bidding competition for the innovation and the profits it earns in case it does not win the bidding competition) *for the monopolist (V^1) and for the entrant (V^2).*

2. Show that $\Pi^m(\underline{c}) \geq \Pi^d(\overline{c},\underline{c}) + \Pi^d(\underline{c},\overline{c})$ is a sufficient condition for $V^1 \geq V^2$.

3. Consider the case in which the innovation is drastic, $p^m(\underline{c}) < \overline{c}$. If this is the case, what is the value of the innovation for firms 1 and 2?

9 Solutions of exercises

Exercise 1. The expected profit of a firm i is given by $\pi_i = \left(x_i / \sum_{j=1}^{n} x_j\right)\Pi - x_i$. The FOCs are therefore given by the system of equations $\partial\pi_i/\partial x_i = \left(\sum_{k\neq i}^{n} x_k\right)\Pi / \left(\sum_{j=1}^{n} x_j\right)^2 - 1 = 0$. (a) At the symmetric equilibrium $\forall i \in \{1,...,n\}$, $x_i = x$, each firm spends $x^* = (n-1)\Pi/(n^2)$. By substitution, one can find that the expected equilibrium profit is $\pi^* = \Pi/(n^2)$. (The second-order conditions for a maximum are satisfied.) (b) Total expenditure nx^* to obtain the monopoly rent Π is given by $(n-1)\Pi/(n)$, which tends to Π as $n \to \infty$.

Exercise 4. (i) (Cournot) To analyse the case of quantity competition, find first the inverse demand function, which is $p = 1 - Q/S$. A firm i chooses q_i to maximise $\pi_i = [1 - (q_i + q_j)/S]q_i - k$. The FOCs are $d\pi_i/dq_i = 1 - 2q_i/S - q_j/S = 0$. At the symmetric equilibrium $q_i = q_j = q$ the solution will be given by $q^C = S/3$. Equilibrium prices and individual profits will be $p^C = 1/3$ and $\pi^C = S/9 - k > 0$. It is easy to find consumer surplus and total welfare as: $CS^C = 2S/9$ and $W^C = 4S/9 - 2k$.

(ii) (Bertrand) Given that products are homogenous, competition would bring about zero (short-run) profits, which would not allow both firms to cover fixed costs. Therefore at the (long-run) equilibrium in pure strategies only one firm will be active in this market. This firm will choose Q so as to maximise $\pi = Q(1 - Q/S) - k$. From $d\pi/dQ = 0$ it follows that equilibrium output, price and profits will be: $Q^B = S/2$, $p^B = 1/2$ and $\pi^B = S/4 - k > 0$. Consumer surplus and total welfare are: $CS^B = S/8$ and $W^B = 3S/8 - k$.

(iii) (Monopoly or cartel) Firms will set outputs q_1 and q_2 so as to maximise the joint profits $\pi_1 + \pi_2 = [1 - (q_1 + q_2)/S](q_1 + q_2) - 2k$. Clearly, this gives rise to the same solution as in the Bertrand case, but with a duplication in the fixed costs. Consumer surplus and total welfare are: $CS^M = S/8$ and $W^M = 3S/8 - 2k$.

The welfare comparisons are straightforward. If $k \in (0, 5S/72]$, then $W^C \geq W^B$. If $k \in (5S/72, S/9)$, then $W^C < W^B$. It is enough to inspect W^C, W^B, W^M to see that W^M is always the lowest.

Exercise 5 Assume $x > 1 - c^h$. Then at the last (price) stage of the game, the innovating firm will be better off by charging the monopoly price $(1 + c^h - x)/2 < c^h$. It will earn gross profit $\Pi_d = (1 - c^h + x)^2/4$.

At the first stage of the game firms correctly anticipate the profit they would make in the following stage. Firms' payoffs are as follows. If both firms innovate, they will both get $-F$. If none innovates, they will both get 0. If only one firm innovates, the innovating firm will get $\Pi_d - F$ and the other firm 0. This game has the following solutions: if $\Pi_d > F$ the only Nash equilibrium is one where one firm innovates and the other does not. If $\Pi_d \leq F$, then no firm will innovate at the equilibrium.

Comparison between monopoly and duopoly

Either market structure makes no difference (innovations occur under both market structures, or do not occur under either), or an equilibrium with innovation occurs under duopoly but not under monopoly. To see why, consider that $\Pi_d > \Delta$ as $(1 - c^h + x)^2/4 > (1 - c^h + x)^2/4 - (1 - c^h)^2/4$. Therefore, for values of fixed cost of innovation such that $\Pi_d > F > x\left(x + 2\left(1 - c^h\right)\right)/4$, there will be an innovation under duopoly but not under monopoly. For values of fixed cost such that $F \leq x\left(x + 2\left(1 - c^h\right)\right)/4$, an innovation will occur in both market structures. Finally, for $F \geq \Pi_d$, no innovation will be adopted, independently of market structure.

Exercise 6 (1) The monopolist will earn profit $\Pi_m(x) = (1 - c^h + x)^2/4$ in the last stage. Therefore, its optimal investment will be determined by finding the value x_m which maximises $\pi_m(x) = (1 - c^h + x)^2/4 - x^2$. It is easy to check that $x_m = (1 - c^h)/3$.

(2) In the case of duopoly the analysis is similar to the one carried out for exogenous levels of R&D. In particular, there will be only one equilibrium where one firm innovates and the other does not. However, this time we also have to determine whether the optimal investment corresponds to a drastic or non-drastic innovation. Therefore, we have to compute the profit made for the optimal drastic innovation and compare it with the profit made for the optimal non-drastic innovation. For a drastic innovation, the domain is $x \in \left[1 - c^h, c^h\right]$. The problem is then the same as the monopolist's. The profit

function reaches a maximum at $x = (1 - c^h)/3$ and then decreases. However, the maximum is not in the domain, and therefore the maximum under a drastic innovation is attained at $x = (1 - c^h)$. In the case of a non-drastic innovation, the optimal amount of R&D x_{nd} is given by the maximisation of the profit function $\Pi_{nd} = x(1-c^h)-x^2$. Therefore $x_{nd} = (1-c^h)/2$. It is easy to check that a non-drastic innovation gives a higher profit to the innovating firm, and that the equilibrium level of R&D under duopoly is larger than under monopoly.

Solution of exercise 8. (a) Since the equilibrium has to be found by backward induction, first solve the last stage of the game, where firms choose quantities given the number of firms that have entered the market in the previous stage. This problem has already been solved in section 3.5.2.

Since n is a real number, the equilibrium number of firms in the industry will be given by the solution of $\pi^c(n) = 0$, which is given by: $n^c = (1-c)/\sqrt{F} - 1$.[83]

(b) The social planner will choose n to maximise total welfare $W = CS + PS = n(1-c)^2(n+2)/\left[2(n+1)^2\right] - nF$, leading to $n^* = \sqrt[3]{(1-c)^2/F} - 1$.[84]

By comparing the optimal number of firms n^* with the number of firms entering at the free entry equilibrium, n^c, it is clear that there exists excess of entry in the industry.[85]

Solution of exercise 9 Welfare can be written as: $W(q(x), x) = (a - p)nq/2 + n\left(pq - (C-x)q - (g/2)x^2\right)$. Given x, the efficient output will be found by taking $dW/dq = 0$. It can be seen that this gives the per-firm output: $q^{sp}(x) = (a - C + x)/n$. Note that at the first-best firms make zero gross profit, as they price at marginal cost ($p = C - x$). The social planner must cover firms' R&D fixed outlays by using subsidies.

[83] If n were to be treated as a discrete number, the equilibrium number of firms in the industry would be given by the value of n which solves $\pi^c(n) \geq 0$ and $\pi^c(n+1) < 0$. This would make calculations slightly less simple but would not change the qualitative results.

[84] The formulation proposed here is inspired by Tirole (1988: 466, exercise 24). He also points out that the first-best would involve the social planner to make only one firm to produce and sell at $p = c$.

[85] The result that free entry gives rise to a number of firms which is larger than the socially optimum number is not robust. In general, the market equilibrium under free entry might involve either too many or too few firms. See also Tirole (1988: chapter 7).

It is now possible to replace this in the previous expression so as to obtain $W(x)$. By solving the FOCs one gets the individual firm's efficient R&D level as: $x^{sp} = (a - C) / (gn - 1)$, which is the socially optimum level of R&D in the industry.

We can now compare the optimal R&D levels with those which arise from competition. It can be checked that the inequality $x^{sp} < x^c$ is satisfied for $n^2 - 2n - 1 > 0$. Hence, when the number of firms is small ($n = 1$ or $n = 2$), then the market produces too little R&D. When the number of firms is larger ($n = 3, 4, ...$), then there is too much R&D as compared to the social optimum.

Solution of exercise 10. Let us start from the last stage of the game. The case of a cartel among the n firms is equivalent to the case of a monopoly, as the n firms behave as if they were a unique firm. In this case, the total quantity sold in the industry will be given by $Q = (a - C + x)/2$ and the joint profit is: $\Pi = (a - C + x)^2/4 - gx^2/2$.

The first order condition is given by: $(d\pi_i/dx_i)|_{x_i = x} = (a - C + x)/2 - gx = 0$, which corresponds, for $n = 1$, to the expression we obtained in the case of competition in section 4.3.2 (note that the competition effect term is absent here). Solving the FOCs will give the per-firm individual R&D level $x^m = (a - C) / (n (2g - 1))$.

The total R&D carried out in the industry will be $R^m = (a - C) / (2g - 1)$. This coincides with $R^c(n = 1)$. Since R^c is increasing with n, it follows that there will always be more R&D under competition than under a cartel. It can also be checked that under a cartel there is always less R&D than at the efficient (social planner) level.

One could also compare welfare levels. Welfare under a monopoly (or cartel) coincides with the expression $W^c(n = 1)$: $W^m = (a - C)^2 g(3g - 1)/\left(2\left[2g - 1\right]^2\right)$.

I showed in section 4.3.2 that the lowest level attained by welfare under competition is corresponding to the limit case where $n \to \infty$, where W^c tends to $(a - C)^2/2$. It can be showed that for $g > 4$ (which satisfies both second order and stability conditions) W^m is always lower than W^c.

Exercise 14

1. If the monopolist decides to sell, the quantity sold during period 1 is re-offered in period 2 (remember a resale market exists). Hence,

a monopolist who has produced q_1 in period 1, will at stage 2 sell the quantity which maximizes its profit, $\max_{q_2} q_2 (1 - q_1 - q_2)$. Notice that the price of the good at period 2, p_2, is the one for which the total quantity produced in *both* periods $(q_1 + q_2)$ is equal to the quantity demanded, $p_2 = 1 - q_1 - q_2$. Simple algebra shows that $q_2 = (1 - q_1)/2$, $p_2 = (1 - q_1)/2$ and $\Pi_2 = (1 - q_1)^2/4$.

Now, at stage 1, the price consumers will be willing to pay for the good will obviously depend on their expectations of p_2. Assume that $E(p_2) = p_2$. Hence, consumer willingness to pay is given by $p_1 = (1 - q_1) + \delta p_2$, the current value of the good plus what resale would give in discounted terms. But, since $p_2 = (1 - q_1)/2$, by substitution one has that $p_1 = (1 - q_1)\left(1 + \frac{\delta}{2}\right)$. Notice that $p_1 > p_2$. When the monopolist sells in each period, prices fall over time. The monopolist then chooses q_1 so as to maximize his present value of (selling) profits Π_s, that is $\max_{q_1} [q_1 (1 - q_1)(1 + \delta/2) + \delta(1 - q_1)^2/4]$. Some algebra shows that $q_1 = 2/(4 + \delta)$, $p_1 = (2 + \delta)^2 /2(4 + \delta)$ and $\Pi_s = (2 + \delta)^2 /4(4 + \delta)$.

2. If, instead, the monopolist decides to lease, then at each period of time t, $t = 1, 2$, it sets a price such that $\max_{p_t} p_t (1 - p_t)$. This implies that $p_1 = p_2 = 1/2$. Then, it produces $q_1 = 1/2$ and $q_2 = 0$, since it was assumed that there is no depreciation The present value of his (leasing) profits is given by $\Pi_l = 1/4 + \delta(1/4) = (1 + \delta)/4$. As can be easily checked, $\Pi_l > \Pi_s$, the monopolist prefers leasing.

Exercise 16

1. At the second stage of the game, if there is entry then the incumbent (I) and the entrant's (E) prices will respectively be: $p_I = \alpha - \beta(q_I + q_E)$ and $p_E = p_I - s$. Note that here we are using the simplifying assumption that no switching costs of buying from the incumbent exist.[86] If there is no entry, the monopolist will set $q_M = (\alpha - c)/2\beta$ and $p_M = (\alpha + c)/2$. The associated welfare will be: $W_M = 3(\alpha - c)^2 /8\beta$.

[86] Otherwise, we should consider a slightly more complex two-period game where only people who have bought from the incumbent will pay no switching cost, whereas people who have never previously bought the good will have to pay the switching cost if they buy from either the incumbent or the entrant. But a two-period game would open the possibility for the incumbent to choose output in period one in a strategic way. See Klemperer (1987c) for an analysis of such a game.

If there is entry, then: $q_I = (\alpha - c + s)/3\beta$, $q_E = (\alpha - c - 2s)/3\beta$ and $p_I = (\alpha + 2c + s)/3$, $p_E = (\alpha + 2c - 2s)/3$. Associated welfare will be: $W_E = \left(8(\alpha - c)^2 + 11s^2 - 8s(\alpha - c)\right)/18\beta$. At the first stage of the game, entry will be chosen only if $s \leq (\alpha - c)/2$.[87] Indeed, for entry to be profitable it must be that $p_E = ((\alpha + 2c - 2s)/3) \geq c$. This condition can be simplified to $s \leq (\alpha - c)/2$. Therefore, if switching costs are large enough, there will be no entry.

2. It is easy to see that $W_E > W_M$ if $s < 5(\alpha - c)/22$. Therefore, for $s \in [0, 5(\alpha - c)/22]$ entry occurs and it is beneficial, whereas for $s \in (5(\alpha - c)/22, (\alpha - c)/2)$ entry occurs but it is detrimental to welfare. This result is due to the fact that entry has two effects. A positive effect, since it decreases prices and increases allocative efficiency. A negative effect, since consumers who patronise the entrant will have to pay a switching cost they would not incur if they bought from the incumbent.

3. If there are no switching costs, then entry will always occur in this model with no fixed cost. The equilibrium output and prices are given by $q_i = q_e = (\alpha - c)/3\beta$, and $p_i = p_e = (\alpha + 2c)/3$. Welfare is given by $W_e = \left(8(\alpha - c)^2\right)/18\beta$. This is always higher than welfare under entry and switching cost.

Exercise 17

1. Let us look at the last stage of the game and solve backwards. At period 2, a consumer who has previously bought from a firm i can be induced to buy from firm j only if $p_j + s < p_i$. For this to be profitable, a necessary condition is $p_j > c$, or $p_j + s > c + s$. Therefore, it must be $c + s < p_j + s < p_i < R$ (the last inequality to ensure that the market exists). Hence, it must be $c + s < R$, which is ruled out by assumption. It follows that undercutting to increase market share relative to period 1 is not feasible, and each firm will then set the monopoly price, that is the price that equals willingness to pay of consumers. In sum, at period 2, $p_A^2 = p_B^2 = R$ and period 2 profits are given by $\Pi_A^2 = q_A^1 N(R - c)$ and $\Pi_B^2 = q_B^1 N(R - c)$, where q_i^1 is the quantity sold by firm i in period 1.

[87] Notice that $q_E \geq 0$ if and only if $s \leq (\alpha - c)/2$.

What happens in period 1? First of all, notice that in period 2 at equilibrium the price will always pay up to her reservation price, R. Therefore, second period considerations do not affect at all consumers' decisions in period 1. A consumer y will prefer to buy from firm A than B if the former gives her higher utility: $r - p_A^1 - ty \geq r - p_B^1 - t(1-y)$, which can be rewritten as: $y \leq 1/2 + (p_B^1 - p_A^1)/(2t)$. All consumers whose learning costs are included in $[0, y]$ will buy from A, the remainder will buy from B: $q_A^1 = 1/2 + (p_B^1 - p_A^1)/(2t)$; and $q_B^1 = 1/2 + (p_A^1 - p_B^1)/(2t)$.

Therefore, total profits after discounting the second-period profit are: $\Pi_A^2 = [1/2 + (p_B^1 - p_A^1)/(2t)]\,[N(p_A^1 - c) + N(R - c)]$;

and $\Pi_B^2 = [1/2 + (p_A^1 - p_B^1)/(2t)]\,[N(p_B^1 - c) + N(R - c)]$. First-order conditions are: $d\Pi_i^2/dp_i^1 = N\left[1/2 + (p_j^1 - p_i^1)/(2t) - (p_j^1 + R - 2c)/(2t)\right] = 0$ (for $i, j = 1, 2$ and $i \neq j$). Focusing on the symmetric equilibrium we obtain: $p_A^1 = p_B^1 = t - R + 2c$.

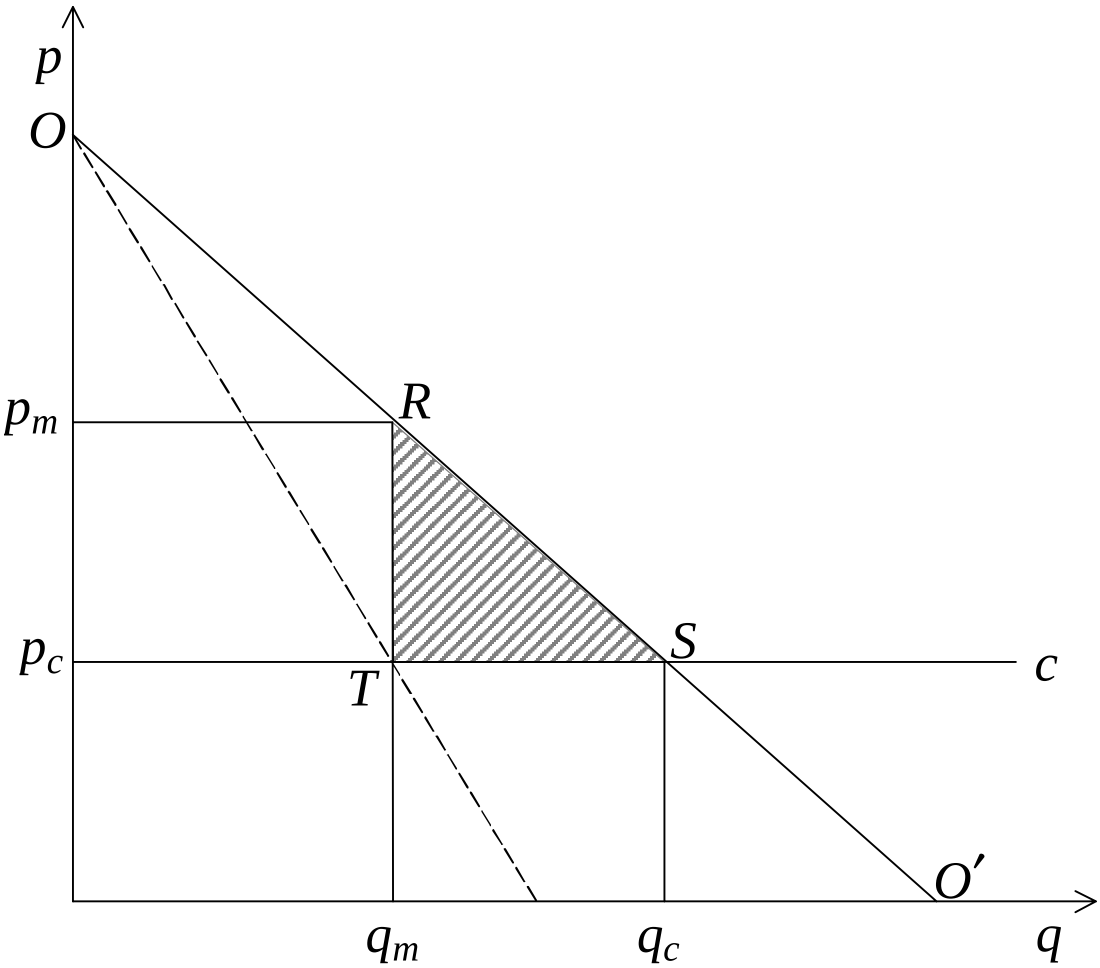
2. If there are no switching costs in the second period, the two products are selling products that are regarded as perfectly homogenous by all consumers. At period 2, therefore, the only possible equilibrium is the usual Bertrand solution, where each firm charges marginal cost: $p_A^2 = p_B^2 = c$. At period 1, consumers's choices will not depend on the second period since they will buy the good at marginal cost anyhow. We therefore have the standard Hotelling-like game being played in the first period. Each firm's profit will be: $\Pi_i^2 = \left[1/2 + (p_j^1 - p_i^1)/(2t)\right] N(p_i^1 - c)$. First-order conditions are: $d\Pi_i^2/dp_i^1 = N\left[1/2 + (p_j^1 - p_i^1)/(2t) - (p_j^1 - c)/(2t)\right] = 0$. At symmetry, we shall have: $p_A^1 = p_B^1 = t + c$. Therefore, in a market with switching costs there will be lower first period prices and higher second period prices than in a market without switching costs.

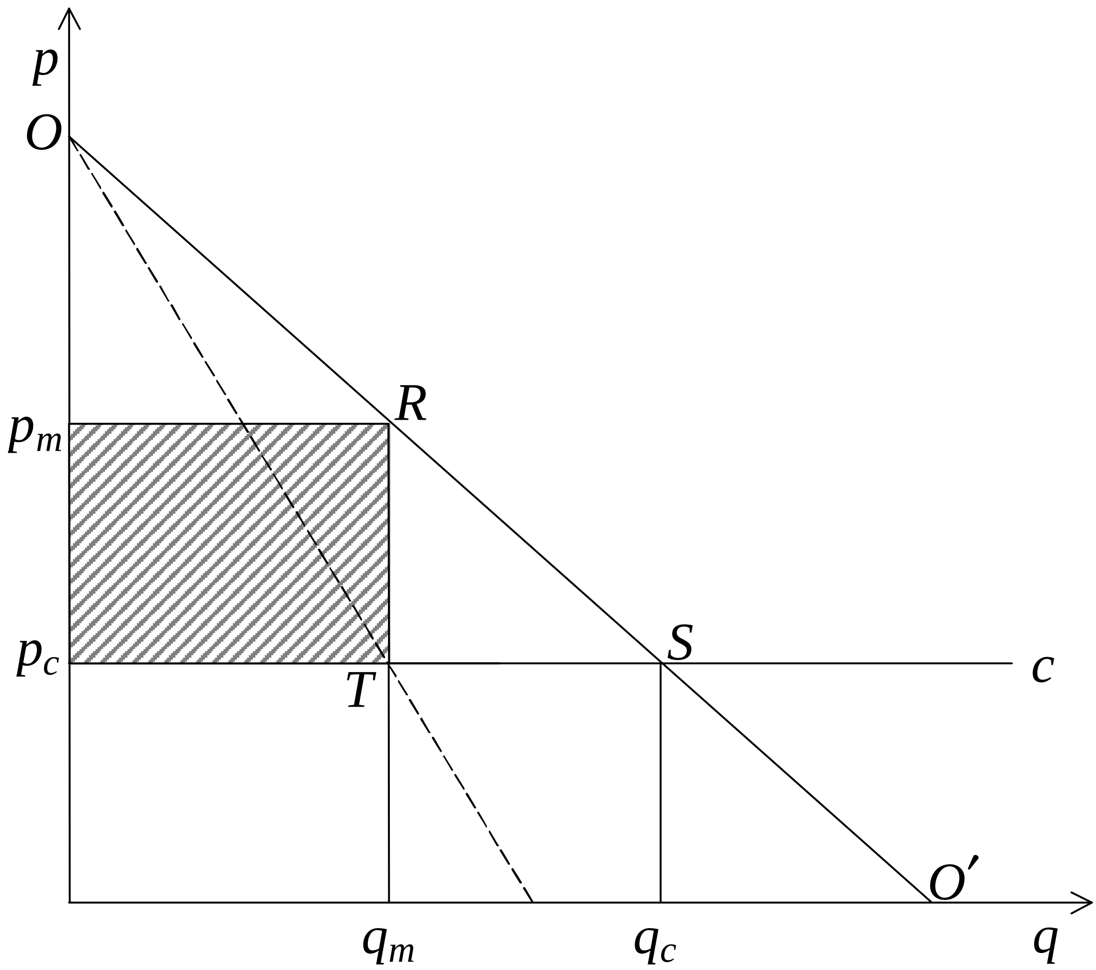
Exercise 18

1. For each firm, the value of the innovation is given by the difference between the profits the firm expects to earn if it wins the bidding competition for the innovation and the profits it expects to earn in case it does not adopt the innovation and the competitor does. Hence, $V^1 = \Pi^m(\underline{c}) - \Pi^d(\overline{c}, \underline{c})$ and $V^2 = \Pi^d(\underline{c}, \overline{c})$.

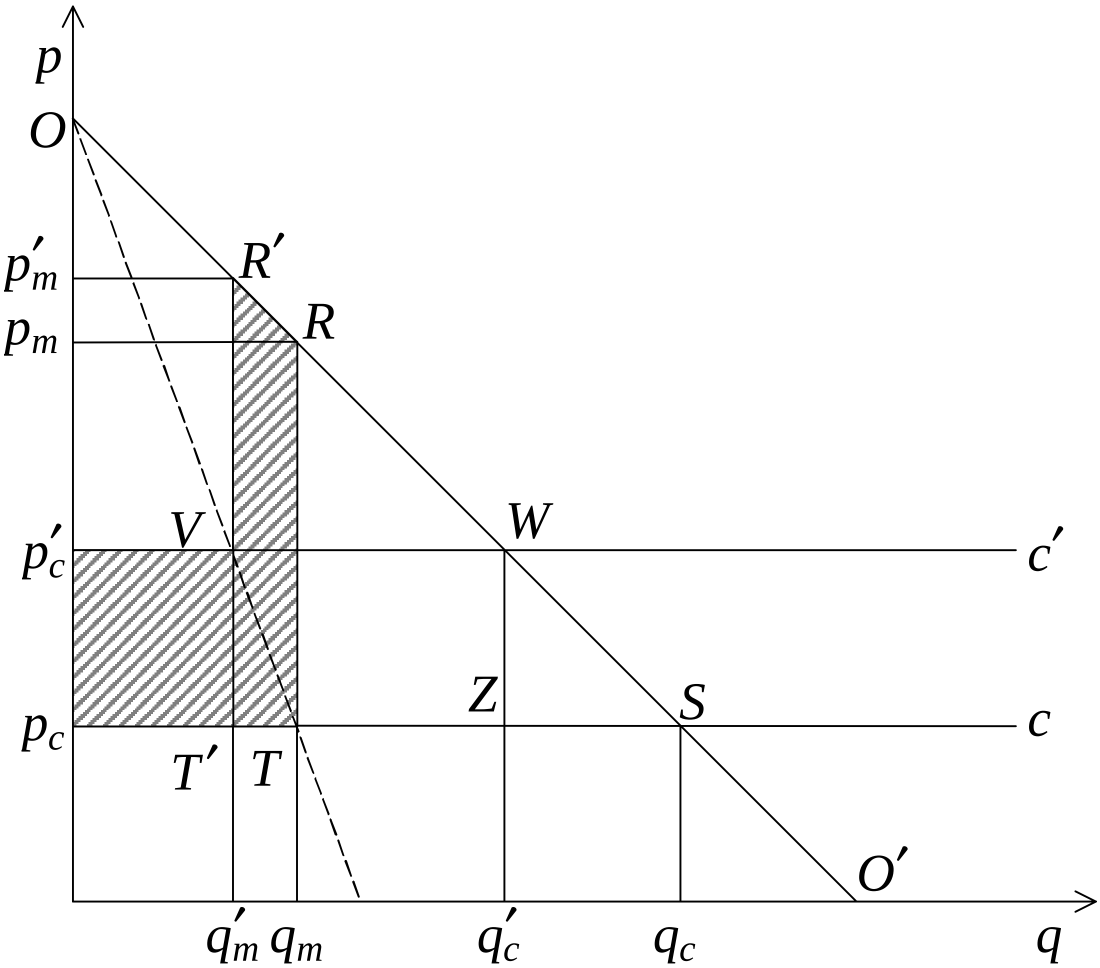
2. Notice that $V^1 - V^2 = \Pi^m(\underline{c}) - \left(\Pi^d(\overline{c}, \underline{c}) + \Pi^d(\underline{c}, \overline{c})\right)$. Hence, if $\Pi^m(\underline{c}) \geq \Pi^d(\overline{c}, \underline{c}) + \Pi^d(\underline{c}, \overline{c})$, then one has that $V^1 \geq V^2$. This property is called the *efficiency effect*, and it is crucial in many models which study whether there is persistence of monopolies: when the industry profits are higher under monopoly than under duopoly, one should expect monopolies to persist.

3. If the innovation is drastic and the entrant wins the bidding competition for the innovation, then he completely eliminates the monopolist from the market. Therefore, $\Pi^d(\overline{c}, \underline{c}) = 0$ and $\Pi^d(\underline{c}, \overline{c}) = \Pi^m(\underline{c})$. This in turn implies that $V^1 = V^2 = \Pi^m(\underline{c})$.

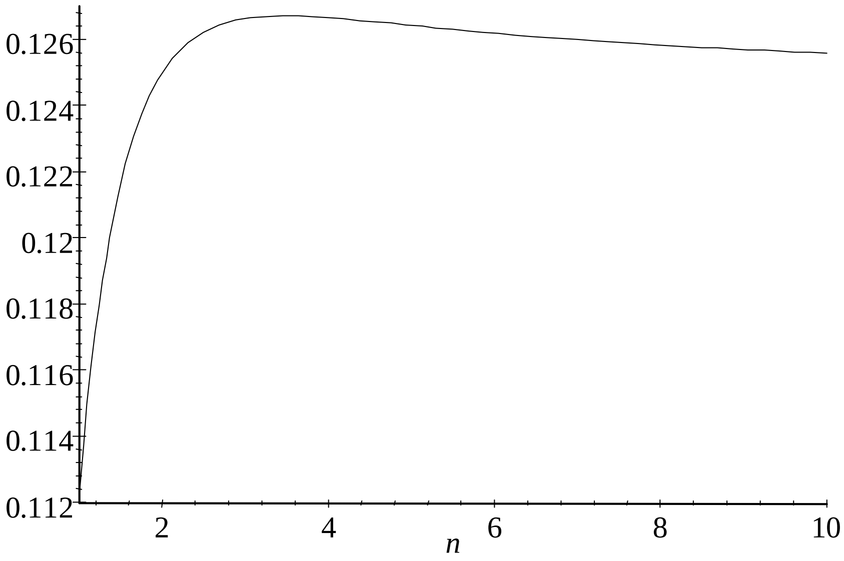

p
O
p_m
R
p_c
S
c
T
O′
q_m
q_c
q



p
O
R
p_m
S
p_c
c
T
O′
q_m
q_c
q



p
O
p′m
pm
R′
R
V
W
p′c
c′
Z
S
pc
c
T′
T
O′
q′m
qm
q′c
qc
q


0.126
0.124
0.122
0.12
0.118
0.116
0.114
0.112
2
4
6
8
10
n