11. How safe is the king’s throne? 
Network externalities on trial

Roberto Pardolesi and Andrea Renda

Network externalities are among the most debated market phenomena of present days. Since Katz and Shapiro published their seminal paper in 1985,¹ many authors have tried to measure the impact of such effects on the so-called ‘neoclassical’ analysis of the market, without reaching any established outcome. Network externalities have been defined as a ubiquitous, pantheistic entity;² they have been taken as supply-side as well as demand-side effects, and have been associated with natural monopoly, essential facilities, learning effects, standardization and tipping.

The remarkable rise of economists’ and lawyers’ attention to network effects is also due to the advent of the so-called ‘post-chicagoan’ approach to antitrust, whose more intense activism stems from the belief that these effects are likely to deviate the market away from its otherwise spontaneous efficiency path. According to this vein of antitrust analysis, network externalities may severely harm competition insofar as they strengthen the barriers protecting incumbent firms from new entrants. Many antitrust decisions of the last decade refer to the existence of network effects as a clue that a certain degree of intervention is needed, even though in most of these cases no solid evidence concerning their effective impact on the aggregate market equilibrium was brought about.³

As we hope to make clear in the following sections, the overwhelming majority of these decisions are based on a vague, nebulous interpretation of the concept here at stake, so as to end up in an embarrassing impasse. Some 15 years after Katz and Shapiro’s first hint, it still remains unclear whether network effects are really as ubiquitous – and by any means related to networks – as they have been credited. Moreover, be they actually widespread or not, it is equally unclear whether their existence alone involves any need for an antitrust inquiry.

A closer analysis of the scenario where the phenomenon is supposed to manifest itself suggests a more cautious approach, aiming to revisit the whole subject from a different, more functional standpoint. Accordingly, this chapter attempts to provide an answer to a few, crucial questions that, in our opinion, still remain partially unsolved. First of all, do network externalities fit their
current definition? Are they truly ubiquitous? Do they generate, and to what extent, undesirable effects? Are network externalities actually related to networks? Eventually, should they be treated as an antitrust issue? Might other remedies, linked to contract law, unfair competition or intellectual property, offer a feasible alternative?

In the first section, we cursorily review the mainstream theory of network externalities. At first glance, they qualify as positive external effects, benefiting both suppliers and final consumers, yet inclined to introduce some level of distortion in the market. The second section contains an attempt to provide a finer tuning of the definition of network effects, together with an analysis of the market conditions that, combined with the network externalities, determine the so-called ‘tipping’ effect and a worrying degree of ‘lock-in’ and ‘path dependency’ on the part of final users. In the third section, we try to understand whether such effects really show up whenever we deal with networks. An even cursory phenomenology reveals that they are lacking in most cases, particularly in the so-called ‘actual’ networks, and that, at any rate, no perverse effect may be associated with their mere existence. In other words, the impact of network externalities on the aggregate welfare requires further analysis, becoming thorough and meaningful only if other peculiar aspects of the markets where they arise are taken into due account.

Hence the fourth section is dedicated to a dynamic welfare analysis of the networks in which final users derive positive external effects from other users’ decision to enter the network. We will argue that these are mainly multidirectional networks, with users dealing with information goods; and we will offer an overall assessment of the peculiar effects that make this kind of market deviate from the typical functioning of neoclassical markets. Finally, the fifth section applies the findings of the preceding sections to the dilemma between the option of laissez faire and an intervention based either on antitrust rules, or on alternative remedies. Some recent cases will be scrutinized on the basis of the theoretical approach provided by the descriptive part of this chapter, thus leading to some prescriptive considerations on how network externalities should be dealt with by regulators.

THE ID OF NETWORK EXTERNALITIES AT FIRST BLUSH

The Mainstream Theory of Network Externalities

According to the prevailing view, if ‘the utility that a user derives from consumption of a good increases with the number of other agents consuming the good’, then network externalities are deemed to be at work. As we will
show, this approach turns out to be exceedingly generic, so that too many heterogeneous phenomena might be associated with it. In this and in the following sections, we try to explain why setting up a finer tuning may prove highly recommendable.

But let us proceed step by step. An externality can be seen as the feedback of an individual activity on other agents in an interactive environment – namely, an unintended consequence of an intended action, generating utilities or disutilities borne by other individuals. In order to qualify for an externality, two elements must be observed: an individual action aiming at the maximization of the agent’s payoff and a collateral effect involving other agents. Thus externalities may be either positive or negative: in both cases, economic efficiency requires that they be completely internalized, letting individual rational decisionmaking lead to an efficient outcome. This appears as one of the most meaningful contributions of the law and economics literature to the understanding of market dynamics. Incidentally, however, it is worth noticing that the internalization of negative externalities has been stressed more frequently than that of positive ones.

Networks are nothing more than the most efficient market structure that can be adopted in many exchange environments. The higher the number of agents involved, the more efficient the choice of a reticular structure. Networking might also be held as the best way of abating transaction costs, in particular distribution and communication costs. Indeed, networks are used both for distribution of a good for final consumption and for communication among end users.

Networks can be divided into actual and virtual. The former are based on a physical infrastructure, such as a railroad track or electrical energy wires. The latter, on the contrary, do not rely on any physical facility, and imply the distribution and exchange of information goods: examples include exchange in cyberspace, information technology markets and so on.

As we have already pointed out, in order to spot an externality we need to observe an individual action that generates (unintended) effects on other agents. In the case of network externalities, the individual action is the decision of an agent to buy the good and therefore enter the network. Network externalities emerge when the value of the good for other agents rises as a consequence of this new entry. Economists measure such a value by observing an agent’s willingness to pay for that good, which depends both on the good’s quality and on its diffusion, namely its network value:

\[ W_A = A_i + A_n \]

where \( W_A \) is the willingness to pay for a good \( A \), yielded by the sum of its so-called ‘intrinsic value’ \( A_i \) and its ‘network value’ \( A_n \). This is a simplified
version of what is usually called hedonic price, in that it distinguishes the components of consumers’ willingness to pay for a good, isolating the relative weight of each of the good’s peculiarities with respect to the final consumption decision of the single agent.\textsuperscript{10}

When a market is characterized by a sufficient degree of network externalities, the widespread diffusion of a good generates a sort of virtuous circle, since users derive a higher utility and therefore tend to value the good more and more, which eventually leads to a larger number of agents entering the network, and so on. The value of a network rises along with the enlargement of its number of users – the so-called ‘installed base’.\textsuperscript{11}

The effect of network externalities on the diffusion of a good is represented as in Figure 11.1. The usual demand curve for a good shifts upwards as a consequence of a higher willingness to pay on the part of agents. Under the same price conditions, the quantity demanded will increase, together with the installed base of users. The extent of the upward shift is that of the higher network value of the good, $\Delta A_n$.

The upward shift in the demand curve highlighted in Figure 11.1 is the result of a demand-driven rise in the value of the good at stake. As we will show in the following sections, these are entirely physiological effects that arise in almost every market exhibiting such characteristics. The new equilibrium will benefit consumers insofar as the price level does not change. Yet,

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure11.1.png}
\caption{The effect of diffusion of good $A$ in a network}
\end{figure}
even if the price for the good rises adapting to the increased demand, the higher value of the network will benefit consumers. The net increase in the value of the network is shared between the standard good producer and the installed base, with the size of the shares depending on the elasticity of the demand curve. In some circumstances, this would be a Pareto-improvement, that is a change that yields no worsening of any agent’s condition in the market, if it were not for the competitors’ lost chance in the race.

One possible shortcoming is that users think twice before leaving the network good in favor (where possible) of a rival, newly entered product $B$, provided that this would be a rational choice only if

$$W_B > W_A \Rightarrow B_i > A_i + A_n;$$

that is, the rival good’s intrinsic value $B_i$ is so much higher than good $A$ that it can compensate its user for the loss of the network value $A_n$, provided that ($B_n = 0$). This may lead to a degree of friction in the market, depending on the relative weight that the network value has over the intrinsic value of the good at stake.

When the network value of a good exhibits a sufficient relative weight, the market may converge towards the selection of a de facto standard good, that dominates for a certain amount of time. Economists call this process tipping; some of them would speak of ‘(excess) inertia’, leading to different market structures, more competitive as the heterogeneity of consumer tastes rises and the critical mass of single networks decreases.

In many cases, more than one network coexist in the same market. The drive towards standardization is seldom strong enough to hinder all potential competitors’ attempts to enter the market. Hence users will choose what they perceive as the best network, and will ‘vote with the feet’ selecting their favorite exchange environment. Given the higher value of larger networks, dominant incumbent networks are more likely to be voted for than small, new entrant ones. When more than one network operate in the market, regulators have to choose between granting competitors access to a single incumbent network or fostering the competition between different networks. The issue comes down to the alternative between ‘internetwork’ and ‘intranetwork’ competition. As we will show in a while, such an alternative becomes mostly important in information-driven markets, where communication goods imply the purchase of primary and complementary goods that build up a ‘system’.

To summarize, the description of the phenomenon leads to some basic considerations:

- network externalities may be viewed as demand-side effects, generated by consumer preferences'
• network effects are mostly positive. When the value of a network to its users depends on the installed base, the demand curve shifts upwards as the number of users increases;
• network externalities may determine a certain degree of tipping, that is a market equilibrium in which a few firms – if not a single actor – gain a substantial share of the whole market, therefore hampering the entry process of would-be competitors. In other words, these externalities work as self-reinforcing peculiarities of consumption. What remains to be assessed is whether this tendency is really so strong that it can lead to a significant distortion in the market, therefore affecting the number of viable competitors.

Interlude: What is Wrong with the Currently Adopted Definition of Network Externalities?

So far, so good. Still, as we reported at the very beginning of this chapter, these externalities are subject to a heated debate and to the j’accuse of many post-Chicago proceedings. Antitrust scholars seem to be concerned with network effects’ alleged contribution to the crystallization of dominant positions in the market. What we have described thus far, however, appears hardly to fit this theoretical scenario. As we will demonstrate in the next section, there is not so much in network externalities as to upset the ordinary matching of supply and demand, and correspondingly to create timeless dominant positions, endangering the process of competition on the merits. As a consequence, either the impact of network externalities is so faint and widespread that it eventually turns out to be of no concern for regulation, and has been starkly overstated by scholars and judges, or there exist other (types of) market effects that, combined with the network ones, pave the way to dangerously anticompetitive outcomes.

More than tipping, lock-in and path-dependency are the major concerns that should alert policymakers. Network externalities, let alone, may just generate an ephemeral tipping: they work exactly as many other frictions that characterize the market in a second best context, and therefore provide an insufficient explanation of the very limited degree of competition observed in some networks. To strengthen this view, one could argue that network effects arise in many non-network markets, such as legal citation techniques in the USA, where one cannot but resort to the West standard, or language learning, insofar as the consumer’s decision concerning which language to learn is highly driven by the number of other agents speaking that language (so that, once an agent has learnt, say, Italian, the value of his investment in learning will be higher the greater the number of Italian-speaking agents and the lesser the number of Italian agents speaking the de facto standard language, English).
As will be clear in the following sections, learning comes close to the core of network externalities. One could also argue – as many did – that in some cases the existence of network externalities determines a standstill situation, in which a new, more efficient solution cannot be selected by the market because of network effects. This reminds us of the famous argument over Qwerty typewriter keyboards as opposed to the Dvorak ones, which allegedly could have led to a higher typing speed, but which did not succeed as a new standard because no user was really willing to switch from the formerly adopted standard to the new one. A similar argument was often raised with respect to the choice of personal computer operating systems and applications, starting from the PC–Macintosh dilemma and ending with the Windows–Linux choice.

It is very important to note that in all the aforementioned cases the existence of network externalities was accompanied by a strong influence of learning investments. In the following sections, this interaction will be dealt with more thoroughly. What we would like to stress at this stage of the analysis is that network effects determine only a limited degree of friction in the market, and at first glance seem to generate neither any true lock-in situation nor any real path dependency in consumption.

A CLOSER LOOK AT NETWORK EFFECTS

Testing the Currently Adopted Definition of Network Externalities: Are There Any Network Effects Inside and Outside Network Industries?

In the last section we alluded to the embarrassment of scholars and judges when coping with what (appears to us a positive, desirable market effect, yet) they tend to treat as the ‘dark side’ of most network monopolies. Let us have a closer look. There are a number of questions that still wait for an answer, concerning where network externalities can actually be found and observed.

As a matter of fact, it is not even clear whether network effects are consistently related to network industries, especially with respect to the so-called ‘actual networks’. We defined those externalities as demand-side effects, arising when the utility that a user derives from consumption of the good rises along with the number of other agents consuming it. Could we apply this rationale to actual networks, such as railroad tracks, electrical energy, gas pipes, television broadcasting? Certainly not. The individual decision to buy the service or not leaves the other users’ condition absolutely unaffected. Among actual networks, only telephones seem to match the usual model, since the intrinsic value of a telephone set is not higher than that of a useless knick-knack, while its network value is undoubtedly predominant.
The same reasoning does not apply to virtual networks, based on no physical infrastructure. The clearest example is that of computer operating systems, whose network value is at least as important as the corresponding intrinsic value.

Not only network effects are missing in several network industries. As long as the current definition is concerned, one could easily ascertain that other phenomena – clearly unrelated to network environments – seem to match such definition, so that the latter ultimately appears too vague. Here are two examples:

**Fashion**: think of trendy, *dernier cri* clothes. Once consumers observe that a certain kind of cloth is more widespread and trendy than others, they will derive a network (social) value from purchasing that good instead of choosing others of the same kind. The wider the spread, the higher will be consumers’ willingness to pay. This leads to very ephemeral dominant positions, showing a periodical tipping in a market for less-than-durable goods and yielding a non-quality-related change in consumers’ willingness to pay for the standard good. This cyclical tipping does not involve any consumer lock-in or path-dependency in consumption.

**Herd behavior**: this occurs when consumers are not able to observe perfectly the quality of a good that is, its experience and credence peculiarities. Consumers may rely on the diffusion of a product as a sort of quality signal, that imperfectly compensates their rational ignorance. As a consequence, the willingness to pay rises, along with the number of other consumers purchasing the good, exactly as happens with network externalities. In this case, like the preceding one, tipping is not accompanied by lock-in or by path-dependency.

**Testing the Currently Adopted Definition of Network Externalities: When do Network Effects Generate Lock-in and Path-dependency?**

In order to find out whether and when network externalities endanger the competitive environment, we need to isolate cases where network effects produce not only tipping, but also some degree of lock-in and path-dependency. In our opinion, this happens only when network effects combine with switching costs, therefore depending upon the kind of good exchanged in the focal market. In this chapter, we focus our attention on learning effects, which we consider to be by and large the most relevant category of switching cost as far as network industries are concerned. Hence it is now time to analyze the interaction of the two phenomena of network and learning effects.

**Demand-side effects: lock-in through learning**

Learning effects are well known to economists, and imply that the value consumers attach to a durable, reusable good rises overtime, as users become acquainted with it. As a consequence, users need to undertake a costly training investment in order to fully enjoy the potential value of the good they
purchased. In other, simpler words, learning effects are positive, demand-side effects that imply a preliminary, costly investment. To be sure, learning how to use a good is costly, and might be viewed as a sunk investment, since conversion to a substitute product may cause the partial or total loss of the effort/investment made to get familiar with the formerly used good. This creates a friction in the market, enhancing the barriers to exit from a technology and therefore partially protecting incumbent firms from the threat of rival technologies. Users find themselves to some extent locked in through learning how to use a product.

Sunk costs associated with the learning process yield the so-called ‘switching costs’. This involves a further gap between agents’ willingness to pay for the de facto standard good and the value of a newly entered product. A hypothetical good \( B \) will be adopted by agents only if

\[ W_B > W_A \Rightarrow B_i > A_i + A_n + C, \]

where \( C \) measures the amount of switching costs agents have to bear when leaving the de facto standard good in favor of the newly entered product \( B \). These costs depend both upon the sunk investment made by agents in order to get familiar with the standard technology and upon the costs required for learning how to exploit the new one. The better and simpler the new technology, the better its chance of becoming the new de facto standard.27

Learning effects are clearly demand side-effects, since they modify the shape of the demand curve in a market. This is shown in Figure 11.2, where the slope of the curve becomes more inelastic. The true believers in the use of a good will progressively value it more and more, whereas those who are not persuaded by its properties will choose to resort to others. The effect on the price level is uncertain, depending on the relative weight of the two factions.

At any rate, the figure clearly shows that the learning process determines an enlargement of consumer surplus. Sunk costs here are necessary to maximize the utility single users derive from the good. The welfare effect is again positive, since no user will bear a sunk investment that she expects will yield an insufficient return. No user will quit the market before recovering at least the value of learning investments.

**Learning effects and network effects combined**

We have already emphasized that network externalities may provide a Pareto-superior change in the market. We also found out that learning effects alone generate desirable welfare consequences, and hence have to be taken as positive effects, though leading to lock-in and path-dependency in consumption.

Yet, when network externalities combine with learning effects, users may find themselves so locked into a dominant technology that it becomes
Figure 11.2  Learning effects, let alone, merely change the slope of the demand curve, increasing the size of consumer surplus

Figure 11.3  The combined effect of learning effects and network effects
irrational to switch to a better product. This might well strengthen the market position of a de facto standard owner. The joint analysis of Figures 11.1 and 11.2 portrays the aggregate impact of the two effects here described. As shown in Figure 11.3, the demand curve changes its slope, becoming more rigid, and shifts upwards, expanding the market and allowing higher profits on the part of the dominant firm.

PHENOMENOLOGY OF NETWORK EFFECTS

After dropping as practically worthless (at least from our standpoint) the traditional dichotomy between actual and virtual networks, it becomes crucial to identify the network industries where positive external effects determine some degree of tipping, which in turn causes lock-in and path dependency. Since the final aim of this chapter is suggesting the most efficient form, if any, of legal intervention, it is absolutely necessary that the markets in need of regulation be properly defined and analyzed.

Unidirectional and Multidirectional Networks

Rather than relying on the dualism between actual and virtual networks, we believe that the likelihood of the surfacing of network externalities depends upon the kind of good that is exchanged in the focal network. This is why we propose to divide network industries into unidirectional and multidirectional. The former are typically distribution or broadcast networks, whose task is to render a certain service or good available to end consumers located in different geographical areas. In this case the network structure is adopted inasmuch as it proves to be the most efficient means of reaching different territorial locations and allows final consumption even by those consumers who are far away from the source of production. Such a structure is therefore exclusively determined on a supply-driven basis, since the selection of the best means to reach final customers is operated by a profit-maximizing producer of a good or service. TV broadcasting, railroad tracks, gas pipes, credit cards and electrical energy wires belong to this kind of market. More generally, almost every market effect observed in unidirectional networks is to be classified as a supply-side effect.

Multidirectional networks, on the contrary, are the only reticular structures in which end agents are not confined to the role of final consumers, but aim to interact and communicate with other agents located along the network. These are more than just networks – they are web structures. Here network externalities are very likely to arise, since the value of the network highly depends upon the number of agents that connect to it, and the network value of the
good is positively correlated with the value of the whole market. The reticular structure is determined by final users’ need to interact, and therefore may be defined as a demand-driven solution. Unlike unilateral ones, multilateral networks are conceived for decentralized communication, not for centralized distribution. The final task of these structures is not allowing final consumption by users located randomly in a geographical area, but ensuring that agents can use the network for communication with other users. In other words, two-way access to the network is what payoff-maximizing agents look for.

Examples of multidirectional networks are telephones (exceptionally, an actual network) and many markets related to cyberspace, such as operative systems, middleware and application software – indeed, virtual networks.

User–User and System–User Interfaces

In our opinion, network effects might raise antitrust or regulatory concerns only in multidirectional networks, conceived for communication. The speed and scope of information exchange between individuals are constantly improved by the fast development of communication technologies. Correspondingly, individuals increasingly demand communication services, instead of choosing more costly means of exchanging information.

Positive externalities arising from the increased size of the network should not be deemed sufficient grounds for a regulatory intervention, insofar as they do not provide for any lock-in or path-dependency in the use of the network. No regulation was intended to hinder rational decisionmaking by self-interested agents. Antitrust rules do not aim to punish successful firms, unless a dominant one abuses its power; and network externalities leave no scope for any reiterated abuse by dominant firms.

As a consequence, multidirectional networks experiencing the mere existence of network effects would not stand for the proposition that special treatment is needed for; antitrust scholars should not be concerned with them more than they are with any other market, at least with regard to demand-side effects. However, when a multidirectional network involves some sort of sunk investments by agents longing to use it, network externalities become a crucial factor for the crystallization of dominant positions, since they enhance users’ barriers to exit from the dominant de facto standard network.

Examples of multidirectional networks where sunk costs do not matter are telephones and fax services, together with other, more traditional communication systems, such as mail. The opposite obtains whenever the network good embodies information goods. These goods regularly involve some sort of learning investments, yet do not always generate network externalities. The information goods used for communication imply both network and learning effects. Information is embodied as software into a hardware device, thus we
call the ensemble a system. Their users need to invest in getting familiar with it; the amount of the investment is called wetware. Hardware, software and wetware are functionally linked and together determine a number of interesting consequences for the welfare and the market analysis.

It is very important to reason in terms of a communication system, that is a whole chain of products whose interaction allows individual use. In the next section, we will explore the issues concerning compatibility and complementarity of goods, whose purchase sometimes involves an additional sunk cost for final users.

As of now, it is worth stressing there are further requirements that have to be satisfied in order to identify relevant network externalities. Firstly, the market must be a communication environment, that is a context in which individuals need to exchange information and choose the most efficient means to achieve this result. We label this feature ‘user–user interface’. Secondly, the interaction between individuals must take place by means of a communication system, consisting of hardware, software and wetware, therefore requiring some degree of learning and expertise on the side of users. We define this requirement as the need for a ‘system–user interface’.

Halfway Remarks

Network externalities were defined as a ubiquitous phenomenon, yet we discovered that, while most networks do not experience such effects, some non-network markets do. If one takes into account that network effects may give rise to lock-in only where a system–user and a user–user interface are indispensable, the phenomenology of dangerous network effects is to focus on information goods, exchanged in the so called ‘knowledge-based industries’.

Since network externalities, let alone, are only positive (though second best and self-reinforcing) effects, regulators should evoke them only when their interaction with other market effects generates peculiar market failures. Moreover, the need for a user–user and a system–user interface clarifies the demand-side nature of the market failure we are analyzing. There is no such effect in those stages of the production chain that do not imply any significant learning investment by users. As a consequence, they would play no significant role in cases like the Intel litigation, though the Pentium processor was considered a sort of essential facility generated, needless to say, by network externalities.

Unidirectional networks, far from experiencing network effects, are created because of supply-side efficiency concerns. Their infrastructure will be at most considered as an essential facility, not a standard. In this case, antitrust and/or regulatory inquiries ought to concentrate on assuring that downstream firms have equal access to the facility, particularly when the latter is proprietary and the incumbent also competes in the downstream markets.
Telecommunications are multidirectional networks. Yet, since no learning effect (that is, a system–user interface) is observed, network externalities do not contribute to the crystallization of the incumbent’s market power. Final users may still experience some sort of switching costs, yet no technological lock-in or path-dependency arises on the demand side. Anyway, even if some significant barriers to exit might be detected in the telecommunication market, they will always derive from a supplier’s strategy that artificially fosters lock-in and ultimately harms consumer surplus in the market. Supply-side interventions, such as mandated interconnection and regulation of access pricing, are the best solution from a policymaker’s viewpoint.

On the contrary, knowledge-based industries are the only networks in which learning and network effects are simultaneously at work. Supply-side policies here can engender the serious risk of decreasing consumer surplus, since the dominant de facto standard is elected by final users amongst the competing ones. There is no real essential facility, in the sense that all hardware and software devices can be duplicated or replaced without incurring unaffordable costs. Invoking the doctrine of essential facilities, as was done in the Intel case, may be appealing from a functional standpoint, but does not seem to be the right approach.\(^{33}\)

In the next section we will focus on the description of knowledge-based industries, in order to understand whether the antitrust concerns raised with respect to network externalities are justified, at least in this field.

**KNOWLEDGE-BASED INDUSTRIES: A WELFARE ANALYSIS**

Network industries do not involve a single market. Indeed, they are chains of interrelated markets, whose links are essential to ensure a proper use of the network good. This peculiar aspect of networks becomes crucial when we turn our attention to knowledge-based industries, because of the characteristics of the goods that are exchanged in this context: information goods. Here network and learning effects interact in a way that can allegedly jeopardize the desirable process of competition on the merits. This section is dedicated to the description of the pathology that otherwise desirable, positive effects may bring into such markets.

**Competing for the Standard: the ‘Gold Rush’ as a Winner-take-all Game**

Network externalities, as we explained in the first section of this chapter, determine a neat tendency towards the selection of a de facto standard good
that dominates the market for a certain time length. Since learning investments generate a degree of path-dependency in users’ choices, the market position of the de facto standard owner will be plausibly stronger than that achieved in other markets. Empirical evidence shows that firms often enjoy substantial extra profits once they conquer the dominant position in such markets.34

As a consequence, rival firms engage in a vigorous competitive race, investing resources in the development of a high quality product. When markets are characterized by a sufficient degree of inertia, this becomes a winner-take-all game, a sort of ‘gold rush’. The higher the expected stake, the stronger the effort of the players.35 At this stage, firms compete to conquer final users’ preferences, therefore prices are likely to fall below average cost. In some extreme cases, empirical evidence shows that products are given away, just because firms seek to spread their products in order to overcome their rivals and achieve the dominant position. At this stage of the game, such a situation benefits consumers, whose surplus becomes larger as competition gets stronger.

Figure 11.4 illustrates this situation, where demand and supply match at point $e^c$, defined as ‘pre-standard equilibrium’. The demand curve is $D$, total output produced is $Q^c$ and the prevailing price $p^c$. Consumer surplus thus equals the area $B$. The pre-standard stage, anyway, yields only a short-run equilibrium. As one firm wins the game, the market will soon become monopolistic and the network will expand as a consequence of network and learning externalities.

Competing for the Standard: after the Gold Rush

If we assume that the final payoff of the winner-take-all game is sufficiently high, the winning player will be granted monopoly profits as long as she keeps controlling the market. The expansion of the network leads to substantial profits and benefits on the part both of the dominant firm and of final users. Point $e^d$ in Figure 11.4 illustrates this new situation.36 The demand curve changes slope, becomes more rigid and shifts upwards as a consequence of combined network and learning effects. In this simplified version of the welfare analysis, consumer surplus is represented by means of the area $A$ and the dead-weight loss arising from monopoly is measured by area $C$.

However, the peculiarities of such markets suggest that the deadweight loss shown in Figure 11.4 is far overestimated. Both demand-side effects, such as sharing or piracy, and supply-side effects, such as indirect network effects, contribute to the widespread diffusion of products among users. In particular, if they cannot afford purchasing an information good, an alternative strategy is to share it. Sharing is a highly common practice in such markets, since individuals may sum their willingness to pay for a product and together decide to
buy and use it. The price level, $p^s$ in Figure 11.4, is hence set at an artificially high level, since dominant firms take sharing into account. If dominant firms manage to discriminate between consumers, for example, through ‘versioning’ of the information good,\(^{37}\) then a separating equilibrium will arise in the market, which maximizes the firm’s profits even though the price is set at too high a level.\(^{38}\)

When sharing is impossible, users may also rely on piracy, which is utterly widespread in knowledge-based industries. Modern technology allows the production of copies whose quality is exactly the same as that of the original product – digital devices are always a matter of numbers, more than materials.\(^{39}\) As a consequence, there is no reason to condemn such monopoly on efficiency grounds. The whole market will be covered by one firm and consumers

---

**Figure 11.4** Welfare dynamic analysis of network and learning effects in markets for information goods
may even be better-off under a monopoly than in a competitive market without tipping. Figure 11.4 shows that area $A$ can be larger than former consumer surplus $B$.

Nonetheless, one may argue that, if rival products were made compatible with the de facto standard, users would enjoy competition in a large network, therefore deriving further utility from using the product. This situation is as likely as having your cake and eating it, since competition at the post-standard stage provides insufficient incentives for competitors to invest in innovation during the pre-standard race. We will anyway deal with the delicate balance between innovation, intellectual property and regulation in the last section of this chapter.

**Information Goods and the King’s Throne: the Pros and Cons of Being a de facto Standard Owner**

The market dynamics described in Figure 11.4 show that final users derive utility from tipping, even if this process leads to a significant reduction in the number of suppliers providing the good. Final users elect the standard good and progressively get familiar with it; hence only final users will decide when to abandon the de facto standard in favor of a new one.

As a matter of fact, the market power acquired by the winning firm in knowledge-based industries differs from what antitrust enforcers are accustomed to regard as dominant position. Peculiar market effects tend to strengthen the winner’s market power, while time and pattern of innovation rapidly erode it. In order to understand whether this situation leads to a market failure, and whether any regulatory path can solve this problem without incurring higher social costs, we still need to answer some crucial questions. Indeed, how safe is the king’s throne? What can the king do to preserve its power in the long run? Will citizens benefit by a more stable power?

Knowledge-based industries often experience a frenetic pattern of innovation, to an extent that the life cycle of the products appears drastically reduced. The winner-take-all game that ends up with the election of the de facto standard is a multiperiod, constantly repeated game; by the time that a firm has become the de facto standard owner, a new game for the next product generation has already begun. As a consequence, firms operating in high-tech industries simultaneously work on products belonging to two or three different generations. The pattern of innovation limits the time length during which the de facto standard owner enjoys her profits.

This might call for no intervention in such markets, since dominant firms need to play and win the repeated game at every stage if they want to preserve their market power and consequently enjoy appealing payoffs, so that competition is assured by the constantly reiterated rounds of the winner-take-all
game. Yet this is only part of the story. Beyond the joint action of learning and network effects, the king’s throne is rendered more solid by indirect, supply-side network effects. These effects generate a substantial degree of market friction, which ultimately grants the standard owner a first-mover advantage at the beginning of the next round of the winner-take-all game. These supply-side effects arise from three different conditions: final users’ technological lock-in, partner firms’ economic dependency and rival products’ incompatibility with the de facto standard. Indeed, an accurate management of customers’ lock-in, a skillful cooperation with partner firms and a vigorous competition with potential rivals appear as the new fundamental chapters in the ‘handbook of the perfect competitor’.42

Primary and Complementary Markets

Information becomes a commodity only when embodied in a physical device, that we defined as an information good.43 These goods typically require some sort of hardware supporting their use. Hence users need to buy and employ a complex set of products, ranging from hardware to software goods, and progressively get familiar with all of them, since only their simultaneous use yields the required communication system. These goods are usually called ‘complementors’.44

Although a network industry is to be viewed as a chain of hardware and software markets, not all the links of such a chain should be bestowed with the same importance. Among the complementors, some are de facto standards and cannot be replaced with competing, compatible products, others are simply chosen out of a range of potentially viable alternatives. We call the formers ‘primary goods’, and the corresponding markets ‘primary markets’ in the network. We also define the latter products as ‘complementary products’, and their markets as ‘complementary markets’ in the network.

The size of a network is determined by the diffusion of its primary good. When final users choose to enter the network, they actually decide to purchase the primary good, which is a de facto standard, and subsequently add their preferred complementary goods. It is with respect to the primary good that users may encounter substantial barriers to exit from the network.

In the case of knowledge-based industries, the de facto standard necessarily corresponds to a software good, since network externalities and learning effects emerge only where both a user–user and a system–user interface are needed. The leading complementor determines the success of the whole network: as far as it keeps being the de facto standard in its market, rival firms will not gain market shares and partner firms – that is, those producing other complementors – will have viable access to the market.

This leads to further considerations. Firstly, the de facto standard owner
may enjoy a substantial contractual power with respect to partner firms that produce complementary products. Since information goods carry privately produced information, protected by intellectual property, dominant firms may enforce their property rights vis à vis competitors as well as all other firms located upstream or downstream in the network chain. As Ayres and Nalebuff put it, the primary good producer will likely take advantage over commercial partners by showing that her BATNA is more appealing than theirs.45 This allows de facto standard owners to extract a quasi rent from each commercial relationship with locked-in partner firms.46 As a consequence, the dominant firm may block all the network markets preventing potential competitors from eroding its market power in the medium to long run. Since commercial relations between firms located in different markets take place through information sharing, that is through the grant of licenses for use of the product, the de facto standard owner will be able to use withdrawal of the license as a threat advantage over every partner. Furthermore, a de facto standard owner will find it relatively easy to expand her power in a complementary market by leveraging her dominant position. One way or the other, weak partner firms could easily be kicked out of their market.

Finally, users wishing to exit the network in favor of a competing new entrant will think twice before doing so. Switching from a standard to a newly entered product might result in forgoing the sunk learning investments, the network value of the currently used good, some of the complementors actually employed together with the primary good, and eventually in bearing further sunk investments in order to get familiar with the new good. There is much more than enough evidence to denote the existence of a friction in the market. If the de facto standard owner is subject to a sufficient competitive pressure, final users will enjoy the attractions of the ‘golden prison’ they were put into. But, if the king’s throne is firm and safe, they will find themselves unwillingly locked into a second-best standard.

The Three Dimensions of Intervention

Though finding out that network externalities, in and of themselves, generate only an ephemeral tipping phenomenon, we stated that, when combined with learning effects, they might give rise to a worrying degree of lock-in and path-dependency on the demand side. This mostly happens when users wish to interact in a reticular structure by means of complex hardware–software systems. Therefore the focus of our analysis was on multidirectional networks, where a system–user and a user–user interface are both needed. As a matter of fact, knowledge-based industries are the only actualization of such a mode.47

When an information good becomes a de facto proprietary standard, its producer is likely to enjoy a conspicuous amount of profits. As a consequence,
the competitive struggle to achieve the dominant position is so vigorous that
the price level might even fall below zero. Correspondingly, dominant firms
exploit all their market power to perpetuate their market position. The coexistence
of direct and indirect network effects, sunk learning investments and
some peculiarities of information goods (such as licensing and the need for
hardware/software complementors) often grants dominant firms some valid
options in this direction. Leveraging, abuse of partners’ economic dependency
and an accurate management of final users’ lock-in all may aptly contribute to
strengthen the leading firm’s dominant position, that is to safeguard the king’s
throne.

Looking at dominant standards from both a horizontal and a vertical
perspective helps overcoming the narrow viewpoint of the single market.
Nonetheless, it is far from easy to say whether, in which direction and to what
extent regulators should intervene in such markets in order to avoid the crys
tallization of market power. Nowhere do issues concerning intellectual prop-
erty, incentives to innovation, compatibility and competition merge and
overlap as in knowledge-based industries.

THE ALTERNATIVE PATHS FOR REGULATING
KNOWLEDGE-BASED INDUSTRIES

Software and the Utopia of Mandatory Open Code

Approaching the conclusion, we try to figure out whether the discipline applic-
able to knowledge-based industries offers any remedy to the potential source
of inefficiencies and market distortions stemming from the existence of a de
facto standard owner. Since final users’ decision to enter a network coincides
with selecting the primary good they want to purchase, complementary
markets entirely depend on such initial choice and therefore are to be dealt
with as aftermarkets. Moreover, as substantial tipping only emerges whenever
both a user–user and a system–user interface are needed, it is very likely that
software (or middleware, general purpose software) becomes the primary
market. We will therefore turn our attention to software markets as the primary
link of the market chains that build up multidirectional, knowledge-based
networks.

Before canvassing the legal technicalities, some light should be shed on
what, at a theoretical level, is often alluded to as the first-best goal to be
pursued. In fact, a seemingly appealing market equilibrium for knowledge-
based industries might be a world where one single standard prevails, yet all
competitors have access to that standard in the form of open code and free use,
or interoperability via various contractual safe-harbor mechanisms, such as de
jure standard setting, patent pooling, and open source and community source licensing is achieved. In this world, consumers would derive a remarkable utility from the existence of a single network, whose value would then be maximized. Learning investments would no longer be viewed as switching costs, since there would be no alternative network to switch to. Moreover, intranetwork competition would drive prices well below the monopoly level.

Unfortunately, the fascination of this approach is largely dispelled by an obvious caveat. Imposing an open standard in such markets might drastically reduce the incentives to creators, resulting in overt inefficiencies. Firms would no longer compete vigorously in the pre-standard stage of the game, since a low expected payoff from even a successful innovation would bring down the initial commitment to innovate. Common sense suggests (despite some remarkable exceptions, deserving closer inspection and in-depth analysis of a different structure of reputational, or other, incentives) that firms do not invest resources in developing a new product to become the pride of its age; however disenchanting this may be, it is expected payoff that fosters competition to achieve the de facto standard. As aptly stressed by an authoritative commentator, ‘[a] firm that manufactures one of the essential components of a network […] would prefer to be the exclusive source of that component rather than be required to disclose the information that would enable competitors to duplicate it. If the component is subject to intellectual-property protection through patent, copyright, or contract (or can be held as a trade secret), then the requisite uniformity is more likely to be achieved by monopoly provision than by standardization’ (meaning some sort of voluntary or mandatory process of convergence).

Despite this warning, the question whether de facto standard owners should be forced to open their source code and let rival firms access the standard keeps being a popular (though controversial) one, which ends up in a debate hinging mainly on the abstract trad-off between internetwork and intranetwork competition.

In our view, intranetwork competition should be the target only when an essential facility, a bottle-neck infrastructure, chokes the development of an acceptable degree of competition. This being the case (which should be ascertained with extremely rigorous caution, since devising the existence of an essential facility should be no short-cut to mere inconveniences of competitors), consumers would certainly benefit from rival firms’ access to the physical, non-duplicable infrastructure that governs the whole remaining market. To put it in the simplest way, imperative economies of scale call for mandatory access to the incumbent’s actual network.

On the contrary, whenever the de facto standard is elected by final users in a repeated winner-take-all game, discouraging innovative efforts by rivals seems hardly appropriate. Scale economies here appear remarkably different
from those observed in actual networks, since they are caused by negligible marginal costs of reproduction, rather than by conspicuous fixed costs. In other words, even a rather small company could cover the whole market: firms in knowledge-based industries are likely to have a very low minimum efficient scale, and virtually no maximum limit, \textsuperscript{56} that is, this is not a natural monopoly. Without competition in the pre-standard stage, neither innovation nor consumers’ welfare would materialize. And such competition is reinforced by the perspective of substantial rewards from licensing the product once it has become a standard.

It should be clear, at this stage of the analysis, that promoting internetwork competition means allocating sufficient property rights to the winning firm; proprietary standards are expected to emerge. In and of itself, this is no evil: as will be seen, the crux is not with the mere existence of protection, but with its extent. Actually, as far as information goods, carrying privately produced information, are concerned, the discipline charged with the task of rewarding its creation is the copyright law. \textsuperscript{57} The innovating firm is thus granted the power to exclude any other agent from (almost any) use of the product for a remarkable length of time. Copyright clearly allows for closed standards, as the spinoff of a set of rules deemed to promote social welfare through monopoly. The predictable consequence is the alluring opportunity, for the firm emerging as dominant, to extract higher tolls from everyone traversing the gateway than could be charged if there were multiple entry points, together with an allegedly lower incentive to innovate. We will revert to this quandary in a while.

\textbf{Antitrust in Knowledge-based Industries: a Bladeless Knife with No Handle?}

Network externalities have been defined as ‘the major antitrust battleground of our contemporary fin de siècle’. \textsuperscript{58} Yet that these market effects dictate antitrust enforcement is a highly debatable question, exacerbated by a recent record studded with striking episodes of clumsiness in coping with the peculiarities of knowledge-based industries. Such impasse mainly derives from the common creed that rules such as those contained in the Sherman Act, or for that matter in the EU Treaty, can be applied without more to information-driven markets and redress their alleged distortions. The underlying idea would seem to be that the basic paradigm continues to be the classic one, so that some sort of corrective manipulation should guarantee a fine revamping of the otherwise declining competitive virtues. But this is precisely the bias we have tried to reject with the preceding analysis. Accordingly, the story should go the other way around: the peculiar features that differentiate knowledge-based industries from the working of mainstream neoclassical markets should
be recognized and duly regarded while evaluating the feasibility and soundness of antitrust intervention. 59

On the one hand, responsibility for the aforementioned impasse may be, at least partially, traced back to confusing definitions of the ‘relevant markets’, mostly missing the fact that the ones we are dealing with are more amorphous than we were accustomed to. 60 Evidence of such a confusion surfaces in each of the resounding cases that have been under antitrust scrutiny over the last few years, such as Kodak, Intel or Microsoft. In the former, Kodak was found to hold a dominant position in the aftermarkets for spare parts and after-sale services on its own produced goods, even though its market share in the primary market was marginal.61 Seemingly, and maybe more paradoxically, Intel was declared to be a monopolist in the market for Intel processors, even if a less tautological approach would have led to a different outcome. 62 Finally, Microsoft, portrayed at once as a ‘near monopolist’ and a ‘fierce competitor’ in the market for Intel-based PC operative systems, has been damned for imposing its Internet Explorer as the standard for browsing.63 Focus on primary markets, therefore treating other complementors as though they were aftermarkets, would have helped to skip ambiguities and misunderstandings.

Yet the very problem lies elsewhere. In the limited range of situations where network externalities actually exert their impact, there will be an inevitable drive toward the creation of a standard, which implies that its owner will be the taking-all winner, at least for some time. This is no way different from the destiny of a firm emerging as the sole actor in the field just because it has ‘outplayed’ the rivals on the merits. The typical antitrust armory, contrasting restrictive practices and monopolization through coercive methods or the like, does not interfere with such a Darwinian process. The trick is represented by the fact that, while the disruptive selection of the participants in the competitive arena seldom comes to the extreme consequence, this is doomed to be the physiological, though temporary, end of the story in knowledge-based industries, for tipping is a necessary precondition to maximization of final users’ welfare and a degree of market power constitutes the reward for developing innovation. There will be a monopolist (though, possibly, ‘fragile’64); and, since under every antitrust sky, becoming a monopolist because of skill, business acumen or luck cannot be a wrong, there is simply nothing that antitrust could (and should) do about it, as long as the dominant firm keeps respecting the qualities that have propitiated its success.65 The magic formula for antitrust to step in is abuse, which might obviously – and often does – materialize, but plays no necessary role in the complex mechanism of network effects leading to a monopoly setting. Therefore, were we to answer the rhetorical question whether it makes sense for antitrust to play a significant role in preventing the entrenchment of monopoly power in knowledge-based industries, we would
say: yes, provided that the intervention aims to condemn the monopolist’s conduct directed at chilling independent and competing innovation, not its mere emerging as a monopolist.

Needless to say, this holds true until the basic provisions of antitrust, that is those relating to firms’ behavior in the market, are considered. Switching to the other ‘soul’ of antitrust, the structural one concerning merger control, would open an entirely different set of evaluations. When asked whether to give leeway to a proposed concentration, the pertinent authorities are to make a prognosis about the future development of the market, in order to avoid growth by acquisition (rather than by physiological expansion) opening a lazy and shorter way to its monopolization. Arguably, this process, with the discretionary power stemming from the possibility of negotiating the terms of a permissible merger, can offer important opportunities to protect both pre- and post-adoption standard competition, as was the case in the **AOL/Netscape** case, and might have happened, on the other side of the Atlantic, with the **Microsoft/Liberty Media/Telewest** transaction (where the EC Commission was inclined to perceive the submitted acquisition by Microsoft of a minority stake in a key UK cable network both as a part of a wider strategy aiming to closely monitor and influence the strategic decisions of software adoption by cable companies, and as a further increase in the chances that the Microsoft package would dominate the future market for set-top box software of the new generation). But the ultimate outcome of the case – which was dropped because the parties withdrew the transaction, though pursuing their goal with the expedient of restructuring Microsoft’s stake so that it would not fall under merger regulation – makes it clear that this kind of control, with its regulatory flavor, is too episodic a way of tackling the issue. Occasional regulation is no promising way of curing a market failure still waiting to be identified as such.

Back to our point. Markets relating to knowledge-based industries should be dealt with according to their peculiarities. If the ultimate goal of antitrust rules is maximizing consumer benefit in the long run, which also means promoting innovation, any chosen pattern of intervention should assume consumers’ long-run interests as its primary issue. And, since tipping is generated by demand-side effects, it seems sound to recommend a cautious approach to supply-side remedies, such as reshaping the markets or breaking up dominant positions by requiring leading firms to open their code.

Even if antitrust turns out to be of little help, there is no reason to raise the white flag and sadly conclude that conduct by the de facto standard owner necessarily, falling in a legal vacuum, skips any scrutiny. To some extent, such behavior may still be effectively addressed by means of other disciplines, whose application, though exhibiting a narrower scope than that of antitrust rules, helps to stigmatize the inefficient features of software markets’ dynamics.
The Need for Intervention outside the Scope of Antitrust: Trojan Horses and Quasi Rents

We have already explained why competing firms should be left free to choose the architecture of their network good, while taking part in the pre-standard winner-take-all game: closed standards do not necessarily yield a market failure.

It is plausible to assume, anyway, that final users would prefer an interoperable product to a closed one, since the former minimizes their switching costs and maximizes the value of a network, increasing their choice of complements in the pre-standard winner-take-all game: closed standards do not necessarily yield a market failure.

On the other hand, while selecting their initial strategy, firms clearly have an incentive to let their product be almost free-ridden, resorting to liberal licensing. This, of course, does not mean that they commit themselves to keeping it freely accessible in the long run, abiding by this tenet also when the good will have become the de facto standard. This sort of ‘Trojan horse’ strategy implies giving away a network good at the pre-standard stage, conquering consumers’ preference (and reliance), winning the game and eventually switching to a closed policy by the strict enforcement of previously disregarded property rights. Once final users have become sufficiently locked-in and the tipping point has been reached, a profit-maximizing strategy implies the exploitation of the huge profits available to the monopolist. Since this policy switching leads to inefficient reliance on the part of consumers, one might claim this is to be considered as an undesirable market failure, which is by no means justified by the need to provide incentives for competing firms.

Switching to a strict enforcement of the closed standard, its owner substantially reduces final users’ choice of primary goods and complements, allowing for a very limited set of alternatives at each stage of the market chain. Final users would certainly suffer from this about-turn. Since they are already locked in, the leading firm will be able to raise prices, at least to a limited extent, without suffering drawbacks in terms of installed base. Admittedly, there is no way to address this problem directly focusing on consumer protection. Somebody has proposed to construe licenses as though they were adhesion contracts, containing the openness of the standard as an implied term, which would play the role of a warranty. But this modified version of ‘copy-left’ seems too tentative a way to fit the needs of a market still looking for identity.

It is worth stressing, however, that a change in the de facto standard owner’s conduct in the market does have other undesirable consequences, particularly as far as vertical relationships are concerned. Final users are not the only agents involved by the switch from a liberal policy to the strict exploitation of the standard. Partner firms are put in exactly the same condition, since they are to forestall the opposite party’s moves and adjust their
expected payoff according to the chance that the latter misbehaves in the long run. They would do better to enter into agreements with ‘open’ players, in order to preserve their chances of dealing with an alternative counterpart and therefore enhance their contractual power along with their BATNA. Suppose the initial story unfolds according to this plot: a later, unexpected ‘ambush’ by the (meanwhile emerged) standard owner, such as the one experienced by the participants to the standard-setting process in the Dell Computer case, with Dell revealing its patent claim only after the involved local bus for VESA (Video Electronics Standard Association) had become highly successful, would overturn their expectations.

Moreover, as was already mentioned, copyright protection allows dominant firms to exploit a substantial degree of bargaining power to the disadvantage of partner firms. This leads to primary goods producers extracting a quasi-rent from their vertical relationships. Indirect network effects and final users’ lock-in bestow upon the primary good producer the possibility of blackmailing the opposite party, by threatening to withdraw its license.

Such paradigms are not reasonably constrained and governed by antitrust, even though the enforcing agencies have tried, from time to time, to expand thus far the reach of their action: yet, in order to achieve this outcome, they always had to configure such conduct as inspired by an attempt to monopolize, and eventually to introduce compulsory licensing by wrongfully applying approaches such as monopoly or essential facilities, which, as we do know by now, absolutely do not fit knowledge-based industries. A less strained alternative for opposing policy switching and quasi-rent extraction may be looked for in the realm of contract law. Primary goods producers usually enter a close net of relational contracts with partner firms at the pre-standard stage of the winner-take-all game. Hence any non-cooperative behavior adopted in the later stage of the game may come under scrutiny as a possible violation of the general requirement of loyalty and good faith in contractual relations. In the case of policy switching, closing the standard in the post-adoption stage of the game arbitrarily reduces the value of the initial investment and causes a sudden enhancement of the barriers to exit from the contract. An analogue evaluation is to be adopted for abrupt license withdrawal (or non-renewal). It is formally true that such behaviors could have been anticipated, and prevented, at the time the contract was originally signed, with the parties operating in a still competitive market. Yet, if it cannot be automatically assumed that a change in practice is unfair, to conclude that holdups and the like are always irrelevant would fly in the face of reality. In a host of circumstances, these strategic behaviors might comprise the reappropriation of opportunities forgone at the time of contract formation, which is commonly held as a violation of the requirement of good faith in the execution of the contract.

Obviously, this remedy turns out to be inapplicable whenever the parties
are not bound by any contractual agreement. Its rationale could nonetheless be implemented by other tools. Some legal systems, such as in Italy, have introduced rules whose application range, though still conceived of as linked to interfirm relational contracting, goes well beyond the boundaries of contractual terms, involving commercial relations at large. These rules address the problem of abuse of either economic dependency, or (according to another definitional strand) a relative dominant position, which exactly fits the pathology we have described as quasi-rent extraction. The comparative advantage of such a remedy stems from the fact that it does not require the existence of a contract between cooperating firms. Consequently, it may also prove useful in coping with the policy switching issue.

But even where one cannot rely on so specific a rule, the underlying suggestion – inducing people to rely on an open policy and then reneging on it when consumers have become locked in is not a merit, it is a wrong – may lead to envisioning an unfair method of competition. Involved firms indeed played the pre-standard stage of the game also relying on the would-be winner’s commitment to leave the good open, whatever the outcome, and did not have the chance to compete on the merits. Disappointment of these reasonable expectations might constitute sufficient ground for invoking the prohibition of unfair practices.

**Abuse of Misuse?**

At this final stage, there is room for an ambitious question, already foreshadowed. Can intervention in software markets preserve both the incentives to innovate and the value-maximizing tipping sources such as network and learning effects? Given the premised analysis, the question echoes the attempt of squaring the circle. But, by now, it is, we hope, clear where the main difficulty lies; and policy suggestions should be shaped according to this learning.

Software industries are mostly resorting to copyright protection. It is worth remembering, however, that copyright law was originally conceived to deal neither with software, nor with tipping or path dependency on the part of final users. Its traditional subject matter were literary and artistic works, where large-scale market power was, to say the least, an unlikely development; its coverage did not involve functional works, whose value inheres in what they do for human beings rather than in what they say or how they appear to human beings. The extension of copyright protection to computer programs, with the obvious difficulties of inventing around an innovative code, represents a dramatic change, which has spawned a host of problems, much more intriguing than those it has contributed to solve. After all, that the same set of rules is applied to the timeless plays of Shakespeare as well as to the ephemeral dominance of spreadsheets such as Lotus 1-2-3 is, let us concede, strange.
enough; their ability to propitiate a satisfactory tradeoff in the new, unexpected setting to which it has been transplanted would be even more surprising. Indeed, this kind of intellectual property right was introduced to protect authors from free-riding. When applied to de facto standard goods in knowledge-based industries, it shelters dominant firms from competitors and generates a threat advantage vis-à-vis partner firms. Network externalities and learning effects simply transform the problem of protecting the author’s work into a means to enhance the barriers to entry in the market – or, alternatively, a barrier to switch away from the current de facto standard: a good reason, according to some commentators, for limiting the extent to which standard owners may profit from such protection to the disadvantage of partners, competitors and final users.

It should be conceded that a more balanced view would reveal plenty of nuances. Copyright enforcement appears at once too pervasive and too fragile to fit the peculiarities of knowledge-based industries. No doubt, it provides for an exceedingly long protection, as attached to products whose life cycle is often a matter of months, not years. On the other side, intellectual property protection, being bound fast to the application of a hard-to-handle criterion such as the idea/expression dichotomy, inevitably leads to an undesirable degree of uncertainty – a strong tipping might be, and actually was, judged to transform an expression into an idea, hence leading to the denial of redress – springing excess litigation between rival firms. Turning again to the opposite view, copyright law does not contemplate protection of improvements on a secured product: should a rival firm develop an improvement on a closed standard product, no incentive to negotiation, such as blocking patents, would be available. Even worse, though beyond the point at hand, the methods of distributing software often establish direct contact with the holder of the right and the ultimate purchaser, so that the former is offered the opportunity of imposing, via contract, restrictions overcoming the faculties conferred by the law. In sum, the cahier de doléances can be easily extended; but motives for denouncing the improper stance of copyright, as applied to the dynamic features of knowledge-based industries, are already consistent.

This is why applying such an out-of-date discipline to a brand new economic stream seems far from a happy choice. Incentives for innovators were desperately needed in order to preserve the vigor of pre-standard innovation and bring down innovation. But too much protection is no clear advantage. The copyright long-lasting umbrella, with the connected power of impeding incremental improvement, was not designed for the protection of functional works of technology: its extension to software has precipitated an in-depth modification of the intellectual property scenario, whose negative by-products are being perceived only as of this time, despite early warnings about the opportunity to devise an alternative sui generis approach.
In this perspective, the intersection of antitrust and intellectual property laws, which has always presented difficult issues and pointed to a border ‘field of dissonance, yet to be harmonized’, might prove even more challenging. This difficulty is highlighted by the uncontrolled expansion of the ‘copyright misuse’ doctrine observed in the last decade, starting with the *Lasercomb* case. By stating that enforcement is excluded whenever copyright is ‘used in a manner violative of the public policy embodied in the grant of a copyright’, the court in *Lasercomb* merely meant to uphold such a measure as a defense to an infringement claim. Yet the repeated and growing application of such defense in later cases ultimately reveals that copyright misuse is becoming something more than a defense, even though, thus far, no court has upheld it as an independent claim. According to many commentators, copyright misuse represents a ‘middle ground’ way towards efficient enforcement of intellectual property rights in software markets. In our opinion, by applying such defense, judges overtly obey a logic of emergency and admit that the tidal wave of network industries has overturned the original scope of intellectual property protection, piercing the veil over its ill-concealed inadequacy. Since this doctrine leads to an enormous degree of discretion on the part of judging courts, the intensive application of copyright misuse may lead software markets towards an even higher degree of uncertainty, with the risk of confusing once and for all the rules of the winner-take-all game and the soundness of incentives to innovate. One could even assert that abusing copyright misuse is in itself a mistake, because of its predictable spillover in other, more orthodox fields, deserving no destabilization; but it is a mistake aiming to redress another mistake, which does not absolve the former, yet invites us not to forget the latter.

**CONCLUDING REMARKS**

Since the antitrust laws, as they stand, are not much concerned with monopoly as such, the most promising move toward a better handling of the problems we have been considering thus far would be, presumably, to reshape the legal protection granted to software, tailoring it according to the real needs perceived in the field. The most promising, we said, but, by the same token, the most unlikely: erasing a 25-year, well-entrenched development seems simply too titanic an effort to gather the required overall consensus. At any rate, this would be the province of legislators.

Nonetheless, if properly pursued, narrow-scope forms of intervention, such as contractual remedies, rules on the abuse of economic dependency and on unfair methods of competition, can contribute to short-circuiting the vicious spiral towards which information-based industries are inclined.
As a consequence, software markets would experience the coexistence of open and closed policies at the preliminary stage of the winner-take-all game. Since copyright protection, leaving aside the question whether conceptually appropriate or not, is fully granted, players would have to decide beforehand about the opportunity to propose an open or a closed product, therefore entering on such a basis into licensing arrangements with partner firm, and final users. Absent the chance to exploit unduly a pre-existing dominant position, no inefficient first-mover advantage would distort the working of the winner-take-all game. In this context, it will be more likely that competition on the merits is reached. As both partner firms and final users prefer open standards, and the former are, to some extent, protected against the risk of Trojan horse strategies, one might bet that an unraveling result obtains, with open standards becoming the take-all winners.

NOTES

2. Network externalities are defined as ‘ubiquitous’ in Kolasky (1999, 577).
3. For what concerns the peculiarities of the post-Chicagoan approach to antitrust, see Hovenkamp (2001). Amongst many others, the most important decisions mentioning network externalities as a cause of anticompetitive distortions in the market are Intergraph v. Intel (see infra, note 62) and U.S. v. Microsoft. In the latter, Judge Jackson stated that the ‘special economics’ of the relevant market – including economies of scale on the supply side and network effects on the demand side – deserved special treatment. On this point, see the skeptical approach adopted by McKenzie (2000).
4. Information is not a commodity in the mainstream economic sense: when it is not embodied in a physical device, it may be treated as a mere precondition of human decisionmaking. Furthermore, we believe that information should not be defined as a public good, since its production is totally decentralized and – at least in most cases – its value is tightly linked to its scarce diffusion. See Pardolesi (1988) and Pardolesi and Motti (1990). Information goods embody privately produced information, and to a certain extent may be treated as commodities, even though these goods preserve some of the peculiarities of information, such as easy sharing and copying: that is, negligible marginal costs of reproduction. For an accurate description of such peculiarities, see Bakos and Brynjolfsson (1999).
5. The definition is given by Katz and Shapiro (1985, 424), and then refined by Lemley and McGowan (1998, 483).
6. The existence and importance of external effects and undesired consequences of intended actions was already theorized in Popper (1967).
7. Indeed, the economic analysis of the impact of external effects on market equilibria can be traced back to Pigou (1920) and was attributed great success during the 1970s. A whole branch of law and economics deals with the internalization of negative externalities such as those arising from pollution or from dangerous activities, therefore calling a comparison between different remedies such as strict liability, insurance and negligence rules. The need to internalize positive externalities has been stressed with less emphasis, if not for issues such as free-riding and public goods. Copyright enforcement may be seen as a means to achieve the internalization of positive externalities arising from innovation.
9. For a similar description of the coexistence of network and intrinsic values in the same good, see also McGowan (1999).

11. According to the so-called ‘Metcalf’s law’, the pattern of growth for a network value is proportional to $n^2 - n$, where $n$ represents the number of users actually connected to the network. See Shapiro and Varian (1999, 224).

12. The extent to which consumers benefit from the upward shift in the demand curve highly depends upon the elasticity of such curve: the steeper it is, the smaller will be the share of total surplus enjoyed by consumers.


14. According to some authors, networks give rise to standardization only when they reach the so-called ‘tipping point’, that is when their size reaches the ‘critical mass’. When tipping is excessively strong, however, it becomes very likely that in the short and medium run inefficient standards prevail; this situation is usually referred to as ‘excess inertia’.

15. Voting with the feet is substantially easier in virtual markets such as cyberspace than in other, non-virtual environments. See also Elkin-Koren and Salzberger (2000). The theory of clubs, anyway, cannot be usefully applied to virtual networks such as the market for operative systems, where learning effects hamper the immediate switching towards the best network. Furthermore, when consumers’ preferences are substantially heterogeneous, network effects will not drive the market towards complete standardization, and as a consequence more than one good will survive.

16. Judge Jackson, for instance, in the Microsoft case, points the finger at the ‘special economics’ of the software market, as characterized by network effects on the demand side and economies of scale on the supply side.

17. While we mention path-dependency, we indeed refer to cases of enhanced path-dependency, since a degree of lock-in by historical events is ever-present in human decisionmaking. Such a hint was rendered explicit by Arthur (1989), but see also earlier works by Hayek (1965) or David (1985), and completes the scenario of individuals’ bounded rationality as developed by the economic literature during last four decades.

18. As a matter of fact, there exist other phenomena whose effect is to enhance the value of consumption though building up barriers to switch to other goods. One such effect is the attachment of an idiosyncratic value to a durable good. Yet all these effects are insufficiently strong to endanger competition on the merits in the long run.

19. For a famous overview of the Qwerty–Dvorak quest and many other anecdotes, and a skeptical view of the possibility that a better standard is driven out of the market by a lower-quality one, see both David (1985) and Liebowitz and Margolis (1990).

20. See supra, note 8.

21. A closer look reveals that consumers may easily be affected by the aggregate values of consumption, yet this only happens through supply-side efficiency-oriented decisions. As an example, you will be more likely to watch your favorite soap opera if many other consumers wish to watch it. The higher the number of potential aficionados, the higher the chance to have it broadcast. But, apart from the chance of consumption, the value of the product will not change along with the number of other agents consuming it. This is the difference between network externalities and indirect network effects, as defined by Lemley and McGowan (1998), which belong to supply-side effects, and emerge as a consequence of the increase in consumers’ demand for a good.

22. See supra, note 9.

23. According to Rubinfeld (1998), ‘While interest in network industries has grown recently because of increasing economic activity involving dynamic industries where there has been substantial innovation and rapid technological change (such as computers and communications), more traditional industries where fads or bandwagon effects may arise (such as designer jeans) are also characterized to some degree by a form of the same phenomenon.’
24. The characteristics that yield the overall quality of a good are usually divided into three categories: search, experience and credence qualities of a good. For a thorough explanation, mostly applied to issues of consumer protection, see Darby and Karni, ‘Free Competition and the Optimal Amount of Fraud’, 16 Journal of Law and Economics, 69 (1973), and Nelson, ‘Information and Consumer Behaviour’, 78 Journal of Political Economy, 311 (1970). Consumers’ rational ignorance may be substantially justified by the impossibility of gathering sufficient information on experience and credence qualities of both the good and the contract terms they sign to purchase it.


26. See Brian Arthur, ‘Competing Technologies, Increasing Returns, and Lock-in by Historical Events’, Economic Journal, 99, March 1989, pp. 116–31. Arthur suggests that lock-in can result when the following are important: (a) fixed costs, (b) learning effects, (c) coordination effects, and (d) adaptive expectations.

27. The importance of past decisions as factors affecting future actions has been recently reaffirmed by the growing branch that applies behavioralism to law and economics. As some authors have recently stressed, the idea of a rational human being as the pivot upon which the whole analysis is turned should be rejected as wholly misleading. A wider scope for describing the many phenomena that lead human decisionmaking away from the modeled efficiency-path is therefore found. See, for example, Korobkin and Ulen (2000).

28. Such distinction echoes that proposed by Daniel Rubinfeld in a speech addressed to the Software Publishers’ Association on March 24 1998. However, Rubinfeld only referred to ‘communication networks’ as opposed to ‘hardware-software (virtual) networks’. Our use of the terms ‘unidirectional’ and ‘multidirectional’ is borrowed from the engineers’ jargon and reflects a different approach, since it aims at isolating those markets where communication and learning are jointly observed. The speech is available at http://www.usdoj.gov/atr/public/speeches/1611.htm.

29. The Metcalfe’s law, as explained supra, in note 11, is applicable only to communication networks, since it links the increase of a network value to the potential interconnections users can realize inside the given network. Each of the n users will be able to communicate with (n – 1) users, therefore the aggregate possible interconnections will be n(n − 1) = n^2 − n.

30. See supra, note 4.


32. The Intel case is probably the best example regarding the danger of an uncontrolled expansion of discretionary power on the part of antitrust enforcers. We will briefly explain the dynamics of the case in the last section of this work. For a thorough analysis, see Picker (1999) and Papciak (1999).

33. See supra, note 32.

34. Richard McKenzie gives a clear example of the abnormal market capitalization some dominant firms enjoy in the so-called ‘new markets’, focusing on Microsoft’s record: ‘Far from being the largest American company in terms of sales or physical plants or employees’, Microsoft’s market value in November 1999 was 17 times its book value, and its market capitalization was five times that of Wal-Mart and twice that of GE. See Mckenzie, 2000, p. 11). Such outstanding values determine the fierce competition observed between Microsoft and its rivals in order to achieve a dominant position in the market.

35. Since the chances of winning the game are not altered, increasing the final payoff means correspondingly raising the expected payoff of each player. Players will therefore be rationally more inclined to invest resources in the development of would-be standard products. They will also consider the possibility of collusion and coalitions, if they believe that joint efforts can substantially improve their chance to win the game.

36. The figure represents a simplified version of what can truly happen in most markets. Since it will be applied to knowledge-based industries, where marginal costs of reproduction are negligible, assuming constant marginal costs under monopoly does not involve an exceeding simplification in the analysis.
‘Versioning’ is a highly common praxis in most markets. The peculiarities of information goods account for an expansion of such marketing strategy. See Shapiro and Varian (1999).

Bakos and Brynjolfsson (1999) show that, when dealing with information goods, firms can extract some share of consumer surplus by charging an artificially high price and exploiting consumers’ sharing. See also Varian (2000 p. 475).

The development of digital technologies now turned every representation of information into 1s and 0s. Once their sequential combination is known, reproduction becomes immediate. See, amongst many others, McKenzie (2000, 22).

Internet operators usually measure time in ‘Internet years’, whose length is seven times less than that of a calendar year.

Of course, such a conclusion does not apply to cases in which a monopolist unlawfully tries to preserve its power extending it to the successive stages of the game, therefore endangering the desirable goal of competition on the merits.

One such handbook is without any doubt that of Brandenburger and Nalebuff (1996).

See supra, note 4.


BATNA stands for Best Alternative To Negotiated Agreement, which represents the threshold under which contracting parties find it efficient to stay inside the commercial relationship. See Ian Ayres and Barry J. Nalebuff (1997, pp. 631ff).

The seminal contribution on quasi-rent extraction is that of Klein et al., (1978), showing that opportunistic behavior that exploits partners’ lock-in inevitably leads to a deadweight loss at intermediate stages of the production chain and to a distortion in the tradeoff between make and buy, that is between vertical integration and market contracting. See also Renda (2000).

We refer to knowledge-based industries for ease of definition. Indeed, not all knowledge-based industries lead to the scenario we depict in this work. The term ‘knowledge-based industry’ therefore stands here for ‘knowledge-based multidirectional networks’.

For an overly enthusiastic assessment of the benefits deriving from open source software development, see Maher (2000) and, in part, Lerner and Tirole (2000); a more balanced view is in Schallopp (2000) and in McGowan (2000).

Needless to say, this caveat does not hold whenever the market exhibits no inclination toward the emergence of a dominant de facto standard (for example, because no single firm controls or owns all the necessary technology IPR (intellectual property rights) pieces of the puzzle; Schallop, 2000, p. 212). Which setting will prevail in a significant number of network computing situations is still unclear.


Kobayashi and Burtis (2000) refer to this tradeoff as to the use–creation dilemma.


Among the most recent examples, the pending antitrust appeal involving Xerox Corporation’s right to refuse to license their patents and copyrights, which has generated significant publicity and amicus brief activity. See In re Independent Serv. Org. Antitrust Litig. CSU, 203 F.3d 1322 (Fed. Cir. 2000) (the amicus briefs filed by, among others, the Intellectual Property Organization, argue that IPR owners have a right to exclude that is not limited by economic markets and that exercising such granted IPR rights cannot in itself violate antitrust laws).


Evans (2000).

On the reward model of IPR, see Schallop (2000).

Kolasky (1999).

‘Only time will tell whether the courts will recognize that new-economy industries have
features that have to be considered for sound antitrust analysis’; (Evans, 2000, p. 72). But see Posner (2000), according to whom ‘there is indeed a problem with the application of antitrust law to the new economy, but that is not a doctrinal problem [. . .]. The real problem lies on the institutional side: the enforcement agencies and the courts do not have adequate technical resources, and do not move fast enough, to cope effectively with a very complex business sector that changes very rapidly’.

60. See McKenzie (2000).


64. ‘For example, in 1999 Cisco had over 80 per cent of the market for the high-speed routers that direct traffic on the Internet, Oracle software managed databases on over 60 per cent of Unix networks, the Palm operating system ran about 80 per cent of the handheld computers in the exploding market for “personal digital assistants,” and Intel had 80 per cent-plus of the PC microprocessor market. And although AOL provides links to the Internet and proprietary content to a mere 25 million consumers, it enjoys the lion’s share of the market for non-business access to the Web’ (Evans, 2000, pp. 72–3).

65.Appearances to the contrary notwithstanding, the antitrust laws are not much concerned with monopoly as such [. . .]. The fact that a monopolist buttressed by network externalities may be hard to dislodge even by a firm with a superior technology has no antitrust significance in itself’ (Posner, 2000).


67. See ‘Merger Control in “New Markets”’, Lexecon Competition Memo (Sept. 1, 2000).

68. This proposition is open to disagreement. See Posner (2000): ‘economic theory and empirical evidence have yet to generate a consensus on whether monopoly is on balance good or bad for innovation’.

69. See Patterson (2000).


71. See Muris (2000, pp. 703ff.)


73. See Art. 9 of Public Law n. 192/1998.

74. This is true even for so-called ‘open-source’ software, which is not produced in a true commons, since property rights are held in reserve to discipline violations of community norms. See McGowan (2000). On the other hand, a significant trend toward resorting back to patent protection is being registered. The implications of such a trend are canvassed by Cohen and Lemley, ‘Patent Scope and Innovation in the software Industry’, 89 California Law Review 1 (2001).

75. See Karjala (2000).

76. Image Technical, at 1217.

77. Lasercomb America, Inc. v. Reynolds, 911 F. 2d 970 (4th Cir. 1990).


BIBLIOGRAPHY


Network externalities on trial


Post-Chicago developments in antitrust law