10. Networks and Network Effects

10.1 Direct Network Benefits
10.2 Systems Competition and Proprietary Interfaces
10.3 Physical Networks: the Internet
10.4 Physical Networks: Cell Phones
10.5 Technical Notes

In chapter 5, we discussed how technology rolls forward in a process of creative
destruction governed, among other things, by intellectual property. Intellectual
property affects competition between successive market incumbents, the average
duration of market incumbency, and – through the incentive to make improvements
– the pace of progress. The market dynamics in which one product replaces another
have been particularly evident in digital technologies such as video games,
text-editing software, Internet browsers, spreadsheets and operating systems.

In this chapter we re-visit the phenomenon of market turnover and market
entrenchment in a context where proprietary products are subject to network effects.
A network is a group of consumers who consume the same good. The network
confers network benefits if the utility it provides to each individual user increases
with the total number of users. An example of a good with network benefits is
text-editing software. The value to each user is greater if he or she can share files
with other users. For example, students can send their homeworks to professors by
e-mail, coauthors can work alternately on the same document file, and editors can
insert their changes directly into authors’ documents. A network good can also be
created by a common standard, such as the convention that cars have brake pedals
on the left and gas pedals on the right. The standard allows drivers to share cars.

These are examples of “direct” network benefits in the sense that the number
of other users affects a user’s utility function directly. “Indirect” network effects
arise when users care about some feature of the network good that is likely to be enhanced in a large network, such as the number of applications developed for an operating system, but do not care directly about the number of users.

The interplay among network effects, intellectual property, and innovation was central in the 1998 lawsuit, U.S. v. Microsoft, brought by the Department of Justice (DOJ). So far, this case and its kin brought in Europe and by individual U.S. state governments are the main cases where competition policy has met network effects. Although the 1998 case reveals that network effects have important effects on market outcomes, particularly in conjunction with intellectual property in an innovation context, the case was ultimately decided on relatively narrow factual circumstances. It therefore left unanswered many policy questions that are likely to resurface.

The DOJ charged that Microsoft illegally attempted to monopolize the market for Intel-compatible operating systems and the browser market. The trial court agreed and ordered Microsoft broken up. Microsoft appealed to the D.C. Circuit.

The court begins its analysis by complaining that “there is no consensus among commentators on the question of whether, and to what extent, current monopolization doctrine should be amended to account for competition in technologically dynamic markets characterized by network effects.” In this chapter we consider some of the economic complexities behind the court’s remarks.

The core of the court’s dilemma is that, under the Sherman Act, attempts to acquire or maintain a monopoly are illegal, even though merely being a monopolist is not. This distinction is especially important in markets with intellectual property, since intellectual property already grants a “legal monopoly.” As pointed out in chapter 5, well-designed intellectual property rights can create healthy competition for product improvements, and can lead to sequential market dominance. With sequential dominance, any snapshot of the market will turn up a firm with high market share, but its dominance may be short lived. This was one of Microsoft’s

---

1In addition to these two charges under section 2 of the Sherman Act, the DOJ brought exclusive dealing and tying charges under section 1. These were respectively rejected by the appellate court and remanded for further consideration. For a discussion of the economics related to these claims, see Gilbert and Katz (2001).
replies to the district court’s finding that it had 95 percent of the market for Intel-compatible desktop operating systems.

Microsoft’s other reply was that the market for operating systems should be defined to include all operating systems that run the same “middleware,” and in that market, they did not have a 95 percent share. Middleware is software such as Netscape or Java that sits between the operating system and the application, and makes the application compatible with the operating system. Like Windows Internet Explorer, Netscape lets users access the web. Java was a development tool created for the express purpose of making applications cross-platform compatible. If the market is defined to include all the operating systems that run the same middleware, the Windows operating system does not have a 95 percent share.

The court rejected this argument, holding instead that middleware had not yet achieved the objective of creating a unified market that extended to all operating systems, and that, by trying to kill the cross-platform capabilities of Java and Netscape, Microsoft was trying to avoid that outcome.

In trying to kill these cross-platform capabilities, Microsoft was overzealous and committed various “bad acts” that made it liable under the antitrust laws. These bad acts included putting pressure on vendors not to ship Netscape to buyers. In addition, Microsoft created a so-called polluted Java that ran faster on Windows than the original Java, but was no longer cross-platform compatible. The court would have accepted polluted Java as a legitimate technical choice except for evidence that Microsoft deceived applications developers into thinking that polluted Java still had cross-platform capabilities.

In retrospect, it seems possible that Microsoft could have achieved the same results without resorting to the specific acts that the court relied on to make its judgment. The case therefore does not reach the deeper questions of whether any of this should be illegal as a business strategy. What is proper conduct in network markets? Where do legitimate business strategies stop and illegal attempts to monopolize begin? We will return to the Microsoft case later in this chapter.

The main objective of this chapter is to develop the ideas of direct and
indirect network effects, and how they may cause the market to tip entirely to one product. The need for a common physical infrastructure additionally changes the way network effects do their work. This is illustrated using the examples of the Internet and cell phone standards.

Intellectual property is the policy lens through which we will view the problem of market power in the presence of network effects and exclusive applications. There are three parts of a system that might be protected: the operating system itself, the interface that allows applications to run on the operating system, and the applications. It is natural to assume that the operating system and applications are true “inventions” that need to be protected, but interfaces are not always in the same category.

10.1. Direct Network Benefits

In the “old” microeconomics, consumers’ demands for products depend only on the relative prices of the products, and not, for example, on consumption by other consumers. Thus, each consumer’s demand curve stays fixed if the other consumers make different decisions about what to consume. With fixed demand curves, consumers make their choices based on price alone. If a single price falls, each consumer reduces his or her consumption of the good, moving along the demand curve. In contrast, in a market with direct network benefits, each consumer’s willingness to pay for a good depends on the number of other users. Because of this, consumer demand is described differently than in chapters 2 and 4.

Instead of writing the demand function as the quantity demanded (or number of buyers) depending on price, as in chapters 2 and 4, we will write demand as a willingness-to-pay function. To define the willingness-to-pay function, line up the consumers on the horizontal axis, each \( \theta \) representing a different consumer. The height of the willingness-to-pay function at \( \theta \) is the maximum amount that the consumer \( \theta \) could pay for the good and still be better off than not consuming it. The consumers are ordered such that, for two consumers, \( \theta \) and \( \theta' \), the consumer named \( \theta \) has higher willingness to pay than the consumer named \( \theta' \) if \( \theta < \theta' \). This
Three willingness-to-pay functions of the form \( w(\cdot, n) \) are drawn in figure 10.1. The variable \( n \) is the network size. By including it as a variable in the willingness-to-pay function, we recognize that each consumer’s willingness to pay depends on the network size. This captures the notion that for goods like text editors, each consumer’s willingness to pay for the good depends on how many others are using it. In the figure, the network sizes are ordered \( n_o < n_m < n_c \). Consider the consumer named \( \theta_o \). Looking vertically upward at the two curves above \( \theta_o \), we see that the consumer’s willingness to pay in a network of size \( n_m \) is larger than his or her willingness to pay in a network of size \( n_o \)—that is \( w(\theta_o, n_o) < w(\theta_o, n_m) \). Since the willingness-to-pay curves satisfy \( w(\theta, n_o) < w(\theta, n_m) < w(\theta, n_c) \) for each \( \theta \), the willingness to pay of each consumer increases with the size of the network.

In the old microeconomics, equilibrium entails that each consumer buys the good if his willingness to pay, as indicated by the demand curve, is greater than the price. With network externalities, the same principle applies, but we must ask, which demand curve? The demand curve shifts every time the price changes, because a different price elicits a different number of users, and each user’s
willingness to pay depends on the number of other users.\footnote{Farrell and Klemperer (2001) for a discussion of the antecedents of this idea, and others in this chapter, and more precise discussions of how expectations about the ultimate size of the network may be formed.}

Figure 10.1 is labeled using the convention that there are \( n_o \) users with willingness to pay at least as high as the consumer labeled \( \theta_o \), \( n_m \) users with willingness to pay at least as high as the consumer labeled \( \theta_m \), and \( n_c \) users with willingness to pay at least as high as the consumer labeled \( \theta_c \). Thus, at price \( p_o \), there are \( n_o \) users \( \theta < \theta_o \) who buy the good. It would be inconsistent to read the number of users at price \( p_o \) from the demand curve \( w(\cdot, n_m) \) because there are only \( n_o \) users, not \( n_m \) users, such that \( \theta \leq \theta_o \).

Figure 10.1 thus shows three situations that the monopoly seller could find herself in, selling at three different prices, \( p_o = w(\theta_o, n_o) \), \( p_m = w(\theta_m, n_m) \), \( p_c = w(\theta_c, n_c) \).

Note that neither the price nor the revenue at that price is monotonic in the size of the network. Even though \( \theta_o < \theta_m < \theta_c \) so that \( n_o < n_m < n_c \) it does not hold (as it would in a market without network benefits) that \( p_o > p_m > p_c \). Instead, \( p_m > p_o > p_c \).

The monopolist earns more profit at the combination \( (p_m, n_m) \) than at the combination \( (p_o, n_o) \), and earns zero profit at the combination \( (p_c, n_c) \), since \( p_c \) is equal to marginal cost. The small network \( n_o \) cannot sustain a high price because the network is too small to support a high willingness to pay. On the other hand, a very large network \( n_c \) might also not sustain a high price because, to expand the network that far, consumers with lower and lower willingness to pay must be included. The price is equal to the willingness to pay of the marginal consumer. As the network expands, the increasing network benefits might not outweigh the fall in willingness to pay of the marginal consumer. If so, there will be some intermediate size network that maximizes the revenue. In figure 10.1, the most profitable size is \( n_m \).

We can use figure 10.1 to understand three focal sizes of the network: the
socially efficient size, the size that a monopolist supplier would prefer, and the size
that a competitive market would support. These arguments are given formally in
technical note 10.5.1. Figure 10.1 shows that the optimal-size network may be
smaller than the whole potential user group, and that the networks produced by
both a competitive market and a monopolist are likely to be inefficiently small.
However, if the network effects are strong enough (not shown in figure 10.1), a
monopolist seller and a competitive market would both include the whole potential
user group, and that would be optimal.

If the standard that defines this network market is proprietary—for example,
text-editing software or an operating system—the seller can choose the price. If the
standard is an “open” standard not owned by anyone, firms will enter in response to
positive profit and depress the price to marginal cost, $p_c$. The locations of the brake
and gas pedals are an example of an open standard. All car manufacturers adhere
to the standard, but if it were for some reason proprietary (patent law has not yet
gone this far), then one could imagine that manufacturers would have to pay for
locating their pedals in that way. Competing standards might develop.

In a competitive market, an open standard will lead to marginal-cost
pricing—that is, $p_c$ in figure 10.1. However, even marginal-cost pricing will lead to a
network that is inefficiently small. The optimal usage may be represented by a
marginal user such as $\theta^*$ in figure 10.1. This is because the network benefits are an
uncompensated externality. Pushing the usage beyond $n_c$ creates benefits for the
other members of the network. Since the marginal users do not account for the
external benefits, their decision whether to join the network at the marginal-cost
price is socially too conservative. To put this another way, the other members of the
network would pay the marginal entrant to join, but have no institutional way to do
so.

A proprietor will prefer an even smaller network—for example, the
combination $(p_m, n_m)$. This is the standard problem that arises as a consequence of
monopoly: in order to maximize profit, the proprietor must inefficiently restrict
usage. However, in network markets, the proprietor has an additional problem.
How can the price and network combination \((p_m, n_m)\) be guaranteed? This is by no means easy, and depends on where the proprietor starts. For example, if the proprietor is an entrant trying to introduce a new standard, she will have to build the market from nothing, perhaps luring customers away from other networks.

Table 5.1 in chapter 5 shows how software products have sequentially invaded the market, driving out previous products as innovators made improvements. That kind of tippiness – the market tipping from one product to the next – is also (perhaps especially) a feature of network markets. The inertia created by network effects can work in both directions: it can foster market conservatism or lead to avalanches of adoption.

Adoptions in network markets are likely to be driven by users’ expectations about prices, other users’ adoptions, and the likelihood of new products entering the market. Entrants’ pricing policies must take account of this. To build a network, the proprietor needs some early adopters. Early adopters of a new network good may fear they will be left stranded if everyone else chooses a different network. This may kill the network before it gets started. To attract the initial adopters, the network owner must convince them that the network will be successful, perhaps offering additional incentives like a price break.

Referring again to figure 10.1, suppose that by random good luck or an introductory offer, the proprietor gets into the market with a small user base, such as \(n_o\) in figure 10.1. Other potential users may be using a rival product or no substitute product at all. Simply raising the price to \(p_m\) is likely to have the opposite effect of that intended: the proprietor may scare customers away instead of attracting them. To attract them, she will have to organize a concerted action so that many join her network at the same time, raising its value to each user. That is probably easiest if she lowers the price in the first instance, even though she eventually wants to raise it.

The customers that the entrant is trying to lure may be of two types: current users of a previous network and neophyte users entering the market for the first time. Even with high expectations of success, it may be difficult to lure customers of the current market incumbent. They have already paid for the previous good,
and goods such as software can be used forever without paying anything further. This advantage is compounded by the fact that the user might have to bear a switching cost to learn how to use the improved product. The previous incumbent is presumably aware of this advantage and may try to maximize it—for example, by ensuring that the new, improved product is incompatible with his own. For example, when Borland tried to break into the Lotus 1-2-3 market with an improved spreadsheet in the late 1980s, it tried to make its product backward compatible with Lotus 1-2-3. Lotus 1-2-3 filed a lawsuit alleging copyright infringement, presumably to preserve a switching advantage.

Thus, even where customers of the previous incumbent are willing to upgrade to an improved product, the entrant may have to offer a deal that is considerably better than the value of the improvement. These market dynamics are slightly different than the ones described in chapter 5, in the sense that switching costs make it harder for an improved product to drive out an entrenched product. Nevertheless, a sufficient improvement may succeed. That said, the incumbent may have more incentive to create improvements than a rival has, which will entrench its market power even further.

All of this calls into question whether an incumbent’s share of a network market is a good test of market power for antitrust purposes. With tippy markets, any snapshot of the market will find some firm with a dominant market share. But sequential monopoly is only a problem for competition policy if the price charged by each sequential monopolist is high. As we have argued, the price is constrained at first by the proprietor’s need to attract users of the previous product, and later by a fear of scaring users into embracing a successor. The same fears will cause the incumbent to keep innovating.

High prices are only one source of inefficiency in network markets. If it is difficult to start an avalanche of buying, an entrant may have trouble entering the

---

3See Farrell and Klemperer (2001) for a more complete discussion.

4See Evans and Schmalensee (2002) for more discussion of this point.
market despite a superior product. Ease of entry depends largely on consumers’
extpectations about the adoptions of others. As a consequence, the market may
coordinate on the wrong standard, or fail to switch to a new standard that is better.\(^5\)
This effect can be especially large where the standard is open, since an open
standard does not have a proprietor who can profit from trying to organize a switch.\(^6\)

In network markets, there are the same trade-offs between incentives to
innovate and protecting consumers from high prices as in any other innovative
environment. Open, or nonproprietary, standards are better for consumers once the
network good exists, since competition will drive profit to zero. This looks good for
consumers, but may also undermine the incentive to improve network goods.
Developer of improved products must be granted some immunity from competition
in order to recover their costs of innovation. This is the same argument for
intellectual property as in other contexts.

Finally, we come to the question of what keeps a network proprietary. So far
we have assumed that the proprietor can prevent rivals from entering with products
that take advantage of a common network. In fact, however, many network
standards are open, so that network goods are competitively supplied. Examples of
open standards include sports equipment, where standards for tennis balls, golf balls,
and so on are set by clubs that sponsor competitions, and the standard placement of
gas pedals and brake pedals in cars. Examples of proprietary network goods are
text-editing software and computer games. In addition to the network goods
themselves, the interfaces that make software run on certain hardware, such as
games on consoles and software on operating systems, may be protected by patents,


\(^6\)See Katz and Shapiro (1994) for examples of inferior standards or products which
became entrenched due to network benefits. Margolis and Liebowitz (1999) agree
that this is possible in theory, but marshal evidence that it has been rare in practice.
They argue from contemporaneous product reviews that in many high-profile instances
where one network won out over another, it was because the winning product was
better. Examples include Excel spread sheets winning out over Lotus 1-2-3 and
Quattro, and VHS winning out over Beta for home video recording. See also Evans
copyrights, or trade secrets. That is the topic of the next section.

Perhaps the most important example of open standards are the protocols of the Internet. TCP/IP governs data transfer among servers, and HTTP/HTML is a common standard for writing websites that are readable by all browsers. The openness of these standards creates network benefits without monopoly power. If the protocols had been proprietary, competing vendors with different standards would have wanted to gain market dominance, and then charge users for access to the Internet. This standards war did not happen. One reason is that development of these protocols was heavily funded by the public sector. If left to the private sector, the protocols would likely have been proprietary. This would have rewarded the inventors, but might also have retarded the Internet revolution by facilitating high prices and creating compatibility problems.

We have so far not considered the possibility that incompatible proprietary goods can be made compatible by conversion (see Farrell and Saloner 1992). For example, the Microsoft Word and WordPerfect text editors started out as incompatible market rivals, but in current versions, users of either text editor can save files in formats readable by the other. Similarly, software files written for the Apple operating system can be converted so that they can be read on the Windows operating system. In both of these examples, the converted user file runs on proprietary software.

Whether conversion requires a license depends on the nature of the intellectual property. If conversion does not require a license, or can be facilitated by some type of reverse engineering, it will undermine the rightholder’s ability to profit from the network, just as if the network good were not protected. If conversion requires a license, then a proprietor can maintain the network monopoly by licensing at a high royalty, so that the market price never falls to marginal cost.

10.2. Systems Competition and Proprietary Interfaces

In contrast to direct network externalities, indirect network externalities arise because of the feedback between applications development and demand. The
stylization is that there is a “platform,” such as a game console or a computer operating system, and applications, such as games or text-editing software, which can only be used on that platform. Typically, there is also an interface that makes the applications compatible with the platform (or not). An important part of market structure depends on whether applications are compatible with multiple platforms, or just one.

The indirect network effect arises because applications developers want to write for popular platforms, and consumers want to buy platforms that give access to many applications. Popularity thus feeds on itself, even if there are no direct benefits to consumers of belonging to a network with many members. As a consequence, the proprietor of a platform can try to tip the market to its platform, by creating many exclusive applications. It then becomes very difficult for a rival to enter without a menu of comparable applications.

Following is a simple economic model that shows why proprietary standards might tip a market to monopoly. Assume that platform owners sell systems with incompatible software. If there are two systems, consumers may intrinsically prefer one to the other, but their preference can be overcome by higher quality (more applications) or lower price. As a consequence, one system may drive the other out of the market.

Assume there are two competing platforms, and that each consumer must choose one platform or none. Each consumer has a preference parameter $\theta$ that designates how much more (or less, if $\theta$ is negative) the consumer is willing to pay for platform 1 than for platform 2. If platform 1 has $a_1$ applications and is sold at price $p_1$, assume that the utility received by a consumer with taste parameter $\theta$ is $\theta + (a_1 - p_1)$. Similarly, the utility received by a consumer of platform 2 is $a_2 - p_2$. The consumer buys platform 1 rather than platform 2 if

$$\theta + (a_1 - p_1) \geq (a_2 - p_2)$$

This relationship makes it clear that, out of the total potential users, the number who use each platform will depend on their relative advantages, $(a_1 - p_1)$ and $(a_2 - p_2)$. If $(a_1 - p_1) > (a_2 - p_2)$ then only those users with a strong preference for
platform 2 (negative $\theta$) will buy platform 2. From this fact, and the distribution of $\theta$, we can derive the numbers of users of the platforms, as they depend on the prices and applications. These are the demand curves shown in figure 10.2, $n_1(p_1; p_2, a_1, a_2)$ and $n_2(p_2; p_1, a_1, a_2)$. They are derived more formally in technical note 10.5.2.

Figure 10.2 shows consumer demand for the two platforms when $a_1 > a_2$. The demand for each platform depends on the price of the other, as well as on the numbers of applications. The numbers of applications $(a_1, a_2)$ are assumed to be fixed, and the firms’ demands, $n_1(p_1; \hat{p}_2, a_1, a_2)$ and $n_2(p_2; \hat{p}_1, a_1, a_2)$, are each evaluated at a fixed price for the other platform.

When the proprietor of platform 1 considers the demand response to a change in his own price or applications, he assumes that the price of platform 2 stays fixed at $\hat{p}_2$, and symmetrically for platform 2.

We now investigate the prices that platform owners will charge, and whether they have incentive to increase the number of applications in order to boost demand. That is, what determines $\hat{p}_1$ and $\hat{p}_2$? In practice, the incentives will depend on the market dynamics, as discussed earlier for direct network benefits. However, insight can be gained by considering a static stylization of the problem in which the firms first commit to their applications $(a_1, a_2)$ and then set their prices $(p_1, p_2)$. 
Figure 10.2 shows an asymmetric situation where firm 1 provides more applications than firm 2. We use this diagram to show that an initial advantage in applications can feed on itself to drive the other firm out of the market. In the two diagrams, the equilibrium prices are \((\hat{p}_1, \hat{p}_2)\). These depend on the applications \((a_1, a_2)\) already in place. With the numbers of applications fixed at \((a_1, a_2)\), firm 1’s equilibrium price \(\hat{p}_1\) is optimal conditional on firm 2’s price, and vice versa. Firm 1’s optimal price \(\hat{p}_1\) is the solution to

\[
\max_{p_1} p_1 n_1(p_1; \hat{p}_2, a_1, a_2)
\]

and symmetrically for firm 2.

In characterizing the prices, we are assuming that the costs of developing the applications and platforms are sunk, and (for simplicity) that the marginal cost of providing access to an additional consumer is zero. In deciding how high to price—in other words, in finding the solution to (10.1)—firm 1 takes account of firm 2’s price \(\hat{p}_2\), and of both firms’ attractiveness in terms of applications, but assumes that they are fixed.

The prices \(\hat{p}_1, \hat{p}_2\) shown in figure 10.2 are equilibrium prices that solve (10.1) and symmetrically for firm 2. In the technical note 10.5.2, the solution is worked out algebraically. The incentives to change price can be seen intuitively by inspecting figure 10.2. In choosing its price, firm 1 (symmetrically, firm 2) takes \((\hat{p}_2, a_1, a_2)\) as fixed, so that its own demand curve is fixed. If firm 1 changes its own price, demand will shift along the demand curve. A marginal price increase will reduce revenue by reducing the number of customers. However, it will also increase revenue by collecting more profit from each customer that remains. When these two effects cancel, the firm has no incentive to either raise or lower price. The optimal price thus has a property familiar from monopoly markets: each firm sets the price where marginal revenue equals marginal cost. This implies that the firm with more applications (higher demand) will charge a higher price, as shown. These are the pricing rules that the firms anticipate when they choose their applications initially.

We can now ask what combinations of applications \((a_1, a_2)\) will be stable in the sense that neither firm has an incentive to increase or decrease its applications,
realizing that such a change will have an impact on both firms’ prices. Figure 10.3 shows how applications and consumer demand feed on each other. The higher demand curve in figure 10.3 represents the increased demand for platform 1 when another application is added—that is when $a_1$ becomes $\hat{a}_1$. If firm 1 then raises the price just enough to keep the net attractiveness of the platform fixed (it increases price to $\tilde{p}_1$), then the number of users of each platform stays fixed and firm 1 collects more revenue in amount equal to the horizontally striped area. Alternatively, firm 1 might keep its price fixed, and thus attract customers from the other platform. This would increase its profit in amount equal to the vertically striped area. Finally, firm 1 could choose a price between these extremes, collecting more revenue from each user, but also attracting some customers from the other platform.

Now compare the profit opportunities available to firm 2 (not shown), which starts with fewer applications and a lower price. If firm 2 adds another application, the profit opportunities will be described in a diagram similar to figure 10.3, but with a lower demand curve, hence lower profit opportunities. This suggests that the firm with more applications has a bigger incentive to increase them still more until the other platform finally drops out of the market.

Of course this story does not account for the fact that both firms will change
their prices in response to either firm’s change in applications. Nevertheless, it suggests correctly that the asymmetric situation described in figure 10.2 will not be an equilibrium. The equilibrium described here is solved in section 10.5.2. There are only two types of equilibria: the firms can share the market equally, or one firm can drop out entirely. However, the equal sharing solution is dynamically unstable for the reasons shown in figure 10.3. In fact, this argument may explain why Microsoft was so keen to diminish the cross-platform capabilities of Netscape and Java. Once it was in the dominant market position, it could entrench that position by creating capabilities and applications exclusive to Microsoft.

The fact that closed standards may tip the market to a single platform, so that one proprietor ends up with an “applications barrier to entry” like that in the Microsoft case, is only the beginning of the story. There are at least three subsidiary questions to address: (1) From society’s perspective, is tipping to a single standard necessarily a bad outcome? (2) From a business perspective, when do firms prefer closed interfaces, and when do they prefer open interfaces? (3) From society’s perspective, what policy tools are available to achieve open interfaces when that is best?

As to the first question, a single monopoly with a proprietary standard offers significant efficiency advantages. Every application is used by every user. Further, costs are not duplicated in developing similar applications for different platforms. For both reasons, more applications will be developed, and they will be used by each user. This is shown formally in the technical note 10.5.2. In contrast, competition between incompatible systems will lead to duplicated costs, and each system may have fewer applications than the monopoly. The one positive aspect is that duopoly may lead to lower prices.

Given these trade-offs, an open standard might be the best of all possible worlds. With open interfaces, applications need not be under the control of a platform owner. Instead, platforms and applications can be priced and sold separately. Prices for each piece of the system will be determined by the breadth of intellectual property rights rather than by the market entrenchment of a dominant
standard due to network effects. Of course, this still does not make network benefits irrelevant. If there are huge network benefits to using a single application, then openness may only transfer the market power from the platform owner to the application owner. The application will be used on all platforms. The main drawback is practical: open standards can be hard to achieve if intellectual property is available for interfaces.

Turning to the second question, whether standards are open or closed is generally a business decision rather than a policy choice. What is the best business strategy? Will proprietors always choose closed interfaces? Referring to the Microsoft case, the cross-platform capabilities created by Netscape and Java served as open standards, since they allowed the same applications to run on Windows and other operating systems. But Microsoft seems to have flip-flopped on whether cross-platform compatibility is a good idea. At some point, Microsoft must have consented to making Netscape and Java compatible with the Windows environment, since Microsoft’s applications programming interfaces (APIs) are not public. On the other hand, the 1998 case is largely about Microsoft’s later attempts to get rid of these capabilities. Apparently its business strategy changed, perhaps in making the transition from young upstart entrant to dominant market incumbent.

The ambiguities of the best business strategy can also be seen in the 1992 case, Sega v. Accolade. Sega was an integrated firm that sold game consoles and games, while Accolade was a game developer that supplied games for the IBM platform. Accolade reverse engineered the Sega interface to make its IBM games compatible with the Sega console. One might have thought that Sega would welcome the additional games, which would make its platform more valuable to consumers. Instead, Sega sued Accolade for copyright infringement of the interface. How should we understand this apparent contradiction?

Sega did not necessarily bring the lawsuit because it thought it would be better off with fewer applications. Rather, it wanted to keep control of its interface. If Accolade needed a license to write for Sega’s console, Sega could demand exclusivity. It is true that adapting the IBM games to Sega would dilute any
advantage that its rival IBM might have, but making games exclusive to Sega would improve its market advantage even further.

The structure of the video-game market has been unstable. This may reflect the incentives to compete for the whole market. However, it also reflects the fact that, in two-sided markets (platforms and applications), there is no reason to think that the market should be driven by one side rather than the other. In the late 1980s and early 1990s, game consoles were provided by integrated vendors, so that games were seldom compatible across consoles. The demand for consoles was driven by the demand for compatible applications. More recently, games have taken center stage. Today, consoles must compete for game developers’ favors. In this environment, developers may prefer to make their games compatible with all consoles and to sell them separately.

So far, we have looked at how dominant firms manipulate standards. What about entrants? Apple Computer, which was the first commercially successful desktop computer, was introduced in the early 1980s. It was an integrated system, with proprietary interfaces and proprietary applications. The IBM personal computer, based on the DOS operating system, was introduced shortly afterward. IBM faced a serious challenge in trying to penetrate the market for desktop computers. It therefore chose to make its interfaces open so that third-party vendors could provide a rich supply of applications, and thus make the IBM PC competitive with Apple. The tactic also encouraged third party vendors to enter the market, thereby assuring consumers that IBM would not raise prices if the PC succeeded and leave them unsupported if it did not. IBM’s strategy was successful, and in subsequent years the market shares of Apple and the PC reversed. But computer prices also plummeted, which was good for consumers but bad for IBM. Along with applications, the open standards invited entry of clones into the computer hardware market, in direct competition with IBM. Compaq and Dell are among the companies that began manufacturing IBM-compatible desktop computers.

Similar tactics were used in videocassette recorders. Here, the standard referred to certain protocols that allow a videotape to be played in certain players.
There were two competing standards in the early 1980s: Sony’s Betamax and JVC’s VHS. VHS was a second comer and needed a strategy to gain a foothold in the market. It achieved this by making the standard open. Like IBM, its success came at the cost of enabling competition.\footnote{See Katz and Shapiro (1994) for more discussion.}

Thus, although the best system for society as a whole may be open interfaces with appropriate intellectual property for platforms (hardware) and applications (software). The foregoing cases suggest that firms are often able to keep interfaces proprietary, either as trade secrets or with formal intellectual property protection. Furthermore, they often have an incentive to do so, especially if they can tip the market. The silver lining is that rivals may decide that open interfaces are the best strategy to get into the market. If an entrant succeeds, it will have changed the market structure from one of closed interfaces or standards to one of open interfaces or standards.

Finally, policy tools for keeping interfaces open seem sparse and unreliable. A right to reverse engineer them, such as discussed by Samuelson and Scotchmer (2002), is a partial antidote, but not reliable for complex interfaces such as the APIs in Microsoft Windows. It is also not available for patented interfaces. Courts may occasionally demand that interfaces be made public, but this can only be done if there is a serious antitrust violation. A federal district court found that the “bad acts” identified in the Microsoft case did not rise to this threshold. More generally, the dominant firms in network markets may win their market power quite legally. The best that courts may be able to do in these circumstances is to make sure that firms that promise open standards or interfaces in order to enter a market do not change their minds afterward.

10.3. Physical Networks: The Internet

Who owns the Internet? Who sets the prices for all those fiber-optic lines that send packets of information on their way? How do we pay for it? How \textit{should} we pay for it?
A remarkable thing about the Internet is how invisible it is to the end user, as a physical network. In their day, railroads and long-distance telephone service were high-profile undertakings. By contrast, the Internet seems to have simply appeared. In the previous section, we stressed that the Internet is a virtual network stitched together by common protocols of data transfer called TCP/IP, topped by the worldwide web, with common protocols for reading and writing webpages embodied in HTML and related languages. Here we explore the physical side of the Internet. The Internet is mostly a collection of fiber-optic lines carrying data packets, connected by routers and other hardware. In fact, it is a collection of local networks, which can transfer data using standard protocols.

The Internet had its origin in ARPANET in the 1960s and 1970s. ARPANET was a network of computers created by the U.S. Department of Defense’s Advanced Research Projects Agency, now known as the Defense Advanced Research Projects Agency or DARPA. DARPA’s most enduring innovation was the TCP/IP data-transfer protocols that make the Internet possible. In the 1980s, the Internet was taken over by the National Science Foundation, which wanted to extend data-transfer capabilities to universities and other research institutions. NSF eventually divested its interest to the private sector, and the NSFNET grew into the Internet.

Why, exactly, did the data-transfer capabilities of ARPANET create a globally networked world? Partly it was just an extremely good idea whose time had come. It was embraced wholeheartedly by researchers in government laboratories and universities. In addition, when NSF lifted its restrictions against commercial use, private firms began to interconnect and extend the Internet.

The Internet has a lot in common with “natural monopolies” such as an electricity grid. However, there is an important difference. In addition to being a physical network, the Internet is a virtual network of users who use the same data-transfer protocols to transfer files. Dividing an electricity grid into two grids with different users would not impinge very much on efficiency, as long as each user was connected, and each of the parts had enough power generators and users to
smooth supply and demand. But if someone divided the Internet into two networks, either physically or virtually, then the direct network benefits of sharing would be reduced. The same is true of a telephone system. If the physical network is divided, not every pair of users are connected, and the direct network benefits of a unified network are diminished. In the case of the Internet and telephone exchange, a unified network is not only efficient in terms of minimizing the cost of delivering services, like the electricity grid, but it enhances the value to users by letting all of them communicate with each other. It provides direct network benefits. These network benefits strengthen the “naturalness” of monopolies.

In this regard, it is worth pointing out something that did not happen: there was no standards war as to the protocols that would underlie the Internet. The protocols of the Internet and worldwide web were developed at public expense and put in the public domain. Given what turned out to be at stake, that is probably one of the most fortunate accidents in industrial history. Imagine if there had been competing vendors trying to lure users onto different networks with different, incompatible proprietary standards. The importance of this nonproblem will become more apparent when we discuss attempts to establish the third-generation of cell phone technologies, where owners used intellectual property to make conflicting standards a reality.

We stressed in section 10.1 that open standards have the virtue in network markets that entry will depress the price to marginal cost, and that this is a better outcome for society than monopoly and also better than having multiple proprietary networks. However, when the standard can only be used in conjunction with a physical network, the prospects for marginal cost pricing might be undermined by natural monopoly. Even if an open standard avoids a standards war, regulation may still be necessary.

Of course, most people do not conceive of the Internet as a monopoly at all, natural or otherwise. Instead we conceive of it as a disorganized and decentralized collection of entrepreneurs attaching and detaching from each other at will. The information highway has different lanes owned by different companies, and there are
lots of networked cul de sacs throwing out packets of information that romp merrily down the highway, changing lanes as convenient to get there as fast as possible. This does not sound like monopoly.

How, then, is all of this paid for? Who sets the prices, and can they be understood as competitive?

An important feature of natural monopolies is that they cease to be natural monopolies as soon as the original infrastructure reaches capacity, and must be duplicated in order to meet demand. Of course, to engender competition, the duplication must be such that each user has two or more potential suppliers, pricing independently. If each user is still hostage to a single supplier, monopoly power persists. For example, building two electricity grids and dividing the users between them would not get rid of monopoly power. Things might be different if a second firm was allowed to lay transmission lines in the first firm’s conduit, so that it reaches the same users. But the first firm is likely to oppose this, forcing competitors to dig new trenches. Given this alternative, it is probably more sensible to regulate the natural monopolist.

In the case of the Internet, increasing capacity generally means laying more fiber-optic cable. Duplication may be expensive in the same way as digging trenches. However, due to the network benefits of interconnection, it seems even more difficult to engender competition on the Internet than on an electricity grid. Even with capacity owned by different firms and reaching all users, the owners of the duplicative infrastructure must interconnect. Because money will change hands when they carry each other’s traffic, they may not compete vigorously for users. Competition may be dampened by the fact that the infrastructure providers may profit from carrying each other’s traffic.

The problem of interconnection afflicts telecommunications more generally. In 1996, the U.S. passed a new Telecommunications Act, designed in part to promote competition in local telephone markets, and thus to make them self-regulating. However, the Telecommunications Act does not pretend that just because incumbents are forced to interconnect with rivals, there will be competition. Instead
the act sets cost-based “guidelines” for reasonable interconnection fees. (See Benjamin, Lichtman and Shelanski 2001 or Laffont and Tirole 2000 for more complete discussions.)

It is not our purpose here to unravel the arcane ways telecommunications services are priced and regulated, and the economic consequences, but only to point out that when a unified physical network is needed to achieve the network benefits of an open standard, it is not self-evident that we can rely on competition to avoid monopoly pricing. Nevertheless, we close this discussion with a short overview of how Internet services are priced.

The original public investments in infrastructure provided a cushion of support that launched the Internet and avoided the funding problem, at least in the beginning. After the original infrastructure was privatized, telecommunications companies like MCI and Sprint made available for Internet traffic their long-haul fiber-optic telecommunications lines. Local user networks then began to connect to this backbone infrastructure, with Internet service providers paying the backbone providers and charging their users for access. The backbone providers grant reciprocal “peering” privileges to each other, exchanging similar amounts of Internet traffic without charge.

To oversimplify, users give money to service providers, who give money to backbone providers, who trade traffic with each other. (In some cases, service providers give money to small network providers who then pay the backbone provider.) There is certainly competition among service providers competing for customers. However, the prices they can offer are constrained by what they pay to the next level up in the hierarchy. If the top level of the hierarchy were a monopoly, competition at the bottom of the hierarchy would not help very much. At the top, the backbone providers increasingly charge traffic-based fees rather than “peering” (Telegeography, 2000). The hope is that backbone providers will compete for paying users since the Internet usually offers several different paths for going from one point to another. Nevertheless, the economic consequences of these arrangements are anything but clear. At root, this system potentially has the same problem as
telecommunications more generally, which is that there must be fees for interconnection. If there is a point of stricture where potential competitors price jointly, monopoly pricing can infect the entire system. This is still in flux.

In summary, open standards will not lead to marginal cost pricing if the network is a physical as well as a virtual network. Open standards still have virtues – they can avoid standards wars or competing networks – but they will not have all the virtues pointed to in section 10.1. To emphasize the importance of avoiding standards wars, we now turn to a physical network where intellectual property in standards has defined the market.

10.4. Physical Networks: Cell Phones

An example that links direct network effects, physical networks, and intellectual property is the problem of cell phone standards.

Cellphones transmit information from the user to a cell tower, which then usually connects to the local telephone exchange. The local exchange transmits the information by landline. Information can only be transmitted between the cell phone and tower if both pieces of hardware send and receive information in the same way. They must use the same transmission standard (the same way of coding information in radio waves) and the same radio frequency. Firms can only make roaming agreements, which allow one firm’s customer to use another firm’s cell towers, if they share the same transmission technology and radio frequency. The standard currently used outside of the U.S. is called GSM. It is widely cross-licensed among a consortium of equipment manufacturers. Since American cell phones embed many different transmission standards, they are often not compatible with each other or with GSM. Hence, U.S. companies cannot make roaming agreements with many Japanese and European companies, and even with some other American companies.

One would have hoped that the roaming problem would be solved by the new generation of improved transmission technologies currently being deployed, called 3G (third generation). The standard that is technically most efficient, in the sense of maximizing the information that can be transmitted with given bandwidth, is said to
be CDMA2000, developed by the American firm Qualcomm, which holds numerous patents on it. The Europeans, who do not want to be held hostage to a proprietary American standard, have developed a slightly different version of CDMA, called W-CDMA. It uses some Qualcomm patents under license, but adds its own patented variations. If adopted, the proprietary aspects of W-CDMA would create a counterweight against Qualcomm’s strong patent position. W-CDMA does, however, have the virtue of being backward compatible with cell towers equipped with Europe’s earlier, GSM standard. W-CDMA is being promoted by the European equipment manufacturers over the better transmission technology, CDMA2000. Due to the backward compatibility, the Europeans say this is efficient. The Americans say it is strategic.

At this writing, the proprietors are having a full-blown standards war, battling it out at every level. (For a complete treatment of this subject, see Cowhey, Aronson and Richards 2003.) The players include international standards setting bodies, high-level government representatives, and also the bodies that allocate spectrum, since compatibility requires a common transmission frequency. At the heart of the problem is big business and protectionism. Europe has large, successful equipment manufacturers that do not want to lose their market power to American patents. American equipment manufacturers want access to lucrative world markets. Second-generation providers outside the U.S. have an incentive to support W-CDMA, since that standard is compatible with their already installed GSM towers. But third-generation entrants can try to unseat the second-generation incumbents by supporting the better technology, CMDA2000. For this reason, the rivalries and conflicts spill across national boundaries.

If U.S. and Asian companies variously adopt CDMA2000 and W-CDMA and European companies adopt W-CDMA, potential network benefits outside of Europe are likely to be limited. Roaming will either be impossible or cell phones will have to embody several standards and radio bands. How the market will tip, if at all, is not yet clear. Every time a cell phone company is “signed” for one side or the other, there is a big press announcement. Sprint and Verizon have chosen CDMA2000.
AT&T Wireless chose W-CDMA. China Telecom chose CDMA2000 in return for royalty concessions, but other users in China have chosen W-CDMA. In Japan, the incumbent telephone company DoCoMo chose W-CDMA, and the rival entrant is betting on CDMA2000.

Cellphones illustrate the important fact that direct network benefits can be provided by membership in a unified physical network, and that such a network requires adherence to a common standard. If the standard is open, then rival firms can supply equipment compatible with the standard, and users will have the dual benefits of competitive supply and a fully integrated network. This is essentially what Europe achieved outside the U.S. with the second generation GSM standard. In view of what has happened with the third generation, that achievement looks more and more miraculous.

Standards wars are likely to erupt when standards are proprietary. If standards were only costless, arbitrary choices, there would be no need to make them proprietary. However, CDMA2000 is a real invention, which was achieved at real cost. The absence of patent protection might have dissuaded the inventor from investing. Later on, however, proprietary standards can hinder the development of a unified network.

One added complication is that the efficient adoption path is not obvious. W-CDMA has the advantage of using installed transmission towers, whereas CDMA2000 will require more infrastructure. On the other hand, if eventual adoption of the more efficient transmission standard CDMA2000 is inevitable, why not do it sooner rather than later? The standards war may eventually be resolved with cross-licensing in which all equipment and telephone networks are compatible and use the best technology. However, that result has been a long time coming.

10.5. Technical Notes

10.5.1. Direct Network Effects

Suppose that the consumers are indexed by $\theta \in [0, \infty]$, with mass one on each unit interval. Suppose that each consumer receives external benefits in amount $f(n)$
when there are \( n \) consumers in the network. The number of consumers in the network is \( \int_0^n d\theta = n \). Assume that the marginal cost of serving each consumer is \( c \), and that the willingness to pay of a consumer indexed by \( \theta \) is

\[
 w(\theta, n) = \frac{1}{\theta} + f(n) \tag{10.2}
\]

Thus, consumers are ordered by the inverse of their willingness to pay. For any given price, it is the low-\( \theta \) consumers who will join the network, since they have the highest willingnesses to pay.

To ensure that the network good will be provided, assume that for some \( n > 0 \), \( w(n, n) > c \).

The social benefits of a network of size \( n \) are

\[
 S(n) = \int_0^n (w(\theta, n) - c) d\theta
\]

The optimum \( n^* \) either satisfies \( n^* < \infty \), and

\[
 \frac{d^2 S(n)}{dn^2} \bigg|_{n=n^*} < 0
\]

\[
 \frac{dS(n)}{dn} \bigg|_{n=n^*} = \frac{1}{n^*} + f(n^*) - c + n^* f'(n^*) = 0 \tag{10.3}
\]

or it is optimal to include everyone in the network \( (n^* = \infty) \).

Now consider the pricing behavior of a monopolist. Since the monopolist has no direct control over the size of the network (it only controls the price), there may be multiple equilibria. This is a problem, but suppose the monopolist can engage in marketing strategies to choose the size of network, and can sustain a network of size \( n \) by charging a price

\[
 p(n) = w(n, n) = \frac{1}{n} + f(n)
\]

That is, it charges a price equal to the lowest willingness to pay of any user on the network. Its profit for a network of size \( n \) is then

\[
 \Pi(n) = n [p(n) - c]
\]

27
The monopolist’s optimum is either to include everyone and charge a price equal to
\( \lim_{n \to \infty} f(n) \) or to choose a network \( n^* \) that satisfies
\[
\frac{d^2 \Pi(n)}{dn^2} \bigg|_{n=n^*} \leq 0
\]
\[
\frac{d \Pi(n)}{dn} \bigg|_{n=n^*} = \frac{1}{n} + f(n^*) - c + \hat{n} \left[ \frac{-1}{\hat{n}^2} + f'(\hat{n}) \right] = 0 \quad \text{if} \quad \hat{n} < \infty \quad (10.4)
\]

Finally, consider what would happen in a competitive market with a
nonproprietary standard. Zero profit implies that the price must satisfy \( p = c \). If
\( \lim_{n \to \infty} w(n, n) > c \), we assume that everyone will be on the competitive network.
Otherwise, the size of the network is \( n^c \) satisfies (10.5) and
\[
\frac{dw(n, n)}{dn} \bigg|_{n=n^c} \leq 0
\]
\[
w(n^c, n^c) = \frac{1}{n^c} + f(n^c) = c \quad (10.5)
\]

We can now compare the sizes of networks. For some functions \( w \), the
optimal network, the monopoly network, and the competitive network include the
whole customer base—for example, if the network benefits described by \( f(\cdot) \) lead to
large willingness to pay for all consumers. In that case, monopoly and competition
are equally efficient, even though customers pay higher prices to a proprietor than
they would pay to competitive firms using an open standard.

It is more interesting to consider preferences for which, as the network
expands, the willingness to pay of the marginal user eventually falls faster than the
network benefits increase. In that case, there will be discrepancies between the
competitive network, the optimal network, and the monopoly network.

In particular, suppose that \( f(n) = 1 - (1/2n^2) \). Then
\[
w(n, n) = \frac{1}{n} + \left( 1 - \frac{1}{2n^2} \right)
\]
\[
\frac{dS(n)}{dn} = \frac{1}{n} + 1 + \frac{1}{2n^2} - c = 0
\]
\[
\frac{d \Pi(n)}{dn} = 1 + \frac{1}{2n^2} - c = 0
\]
\[
\frac{dS(n)}{dn} = \frac{d\Pi(n)}{dn} + \frac{1}{n}
\]

The optimal-size network \( n^* \) is larger than the monopolists’ network \( \hat{n} \). Further, the competitive network satisfies

\[
w(n^c, n^c) = \frac{1}{n^c} + 1 - \frac{1}{2(n^c)^2} = c
\]

One can see that

\[
\frac{d\Pi(n)}{dn}\big|_{n=n^c} < 0 \quad \text{and} \quad \frac{dS(n)}{dn}\big|_{n=n^c} > 0
\]

Hence the monopolist will prefer a smaller network than the competitive network, and the socially optimal network is larger than the competitive network.

The monopolist will arrange for a smaller network by charging a price higher than the competitive price. Thus, when \( w(n, n) \) is decreasing in \( n \), \( \hat{n} < n^c < n^* \).

The reason that \( n^c < n^* \) is that an expansion of the network from \( n^c \) would create benefits for the inframarginal members of the network which are not captured by the competitive firms. Indeed, they will have to lower their price below \( c \) to attract new customers. From society’s point of view, this is a good deal, since inframarginal members of the network benefit, while the discrepancy between the marginal consumer’s willingness to pay and the cost of serving him or her is tiny.

But since the firms cannot capture the benefits to inframarginal members, they will not expand the network.

### 10.5.2. Multiple Equilibria in Systems Competition

To derive the demand functions, assume that \( \theta \) is distributed uniformly on \([-1, 1]\).

Since consumers can drop out of the market and receive utility 0, impose that \((a_2 - p_2) \geq 0\). Thus, the alternative to buying platform 1 is buying platform 2, and buying platform 2 is no worse than dropping out of the market. The demands for the platforms, shown in figure 10.2, are

\[
n_1(p_1; p_2, a_1, a_2) = \begin{cases} 
0 & \text{if } 1 - (a_2 - p_2) + (a_1 - p_1) < 0 \\
1 - (a_2 - p_2) + (a_1 - p_1) & \text{if } 0 < 1 - (a_2 - p_2) + (a_1 - p_1) < 2 \\
2 & \text{if } 2 < 1 - (a_2 - p_2) + (a_1 - p_1)
\end{cases}
\]
characterize the prices \( \hat{p}_1(a_1, a_2), \hat{p}_2(a_1, a_2) \) that arise in the subsequent Nash equilibrium where firm 1 (symmetrically, firm 2) chooses price to maximize (10.1). The Nash equilibrium prices satisfy \( \hat{p}_2(a_1, a_2) \leq a_2 \) and \( \hat{p}_1(a_1, a_2) \leq 1 + a_1 - (a_2 - \hat{p}_2(a_1, a_2)) \), and the following:

\[
\begin{align*}
-2\hat{p}_1(a_1, a_2) + 1 + a_1 + (\hat{p}_2(a_1, a_2) - a_2) & = 0 \quad \text{if} \quad 0 < \hat{p}_1(a_1, a_2) \leq 1 + a_1 - (a_2 - \hat{p}_2(a_1, a_2)) \\
-2\hat{p}_2(a_1, a_2) + 1 + a_2 + (\hat{p}_1(a_1, a_2) - a_1) & = 0 \quad \text{if} \quad 0 < \hat{p}_2(a_1, a_2) \leq a_2
\end{align*}
\]

Solving these two equations, we can write the equilibrium prices as a function of the pre-specified numbers of applications, \( (a_1, a_2) \):

\[
\begin{align*}
\hat{p}_1(a_1, a_2) & = \begin{cases} 
\frac{1}{2}(3 + (a_1 - a_2)) & \text{if} \quad 3 \geq (a_1 - a_2) \geq -3 \\
\frac{1}{2}(1 + a_1 - a_2) & \text{if} \quad (a_1 - a_2) > 3 \\
0 & \text{if} \quad (a_1 - a_2) < -3
\end{cases} \\
\hat{p}_2(a_1, a_2) & = \begin{cases} 
\frac{1}{2}(3 - (a_1 - a_2)) & \text{if} \quad 3 \geq (a_1 - a_2) \geq -3 \\
\frac{1}{2}(1 + a_2 - a_1) & \text{if} \quad (a_1 - a_2) < -3 \\
0 & \text{if} \quad (a_1 - a_2) > 3
\end{cases}
\end{align*}
\]

Thus \( \hat{p}_1(a_1, a_2) - \hat{p}_2(a_1, a_2) = \frac{2}{3}(a_1 - a_2) \) if \( 3 \geq (a_1 - a_2) \geq -3 \), and the number of customers of firm 1 is \( n_1(p_1; p_2, a_1, a_2) = 1 + (a_1 - a_2) - (\hat{p}_1(a_1, a_2) - \hat{p}_2(a_1, a_2)) = 1 + \frac{1}{3}(a_1 - a_2) \). If \( a_1 - a_2 > 3 \), the firm 1 has all the customers, namely, a mass of two.

We can now define the firms’ profit functions, as a function of \( (a_1, a_2) \). Suppose that there is a pool of applications developers with different talent or efficiency, so that the marginal cost of providing applications is increasing. Assume the total cost of providing \( a \) applications is \( a^2/6 \), regardless of how the applications are split.
between the firms. Assume in addition that each firm pays its proportional share of
the total cost of applications development—for example, firm 1 pays

\[
\frac{1}{6}(a_1 + a_2)^2 \frac{a_1}{a_1 + a_2} = \frac{1}{6}a_1(a_1 + a_2)
\]

Then

\[
\Pi^1(a_1; a_2) = \begin{cases} 
2\left(\frac{1}{2}\right)(1 + a_1 - a_2) - \frac{1}{6}(a_1 + a_2)a_1 & \text{if } a_1 > a_2 + 3 \\
-\frac{2}{5}(a_1 + a_2)a_1 & \text{if } a_1 \leq a_2 - 3 \\
\frac{1}{5}(3 + a_1 - a_2)^2 - \frac{1}{6}(a_1 + a_2)a_1 & \text{if } a_2 + 3 \geq a_1 \geq a_2 - 3
\end{cases}
\]

The profit function of firm 2 is the same, except with the subscripts reversed.

There are two Nash equilibria in the number of applications. There is a
symmetric equilibrium with two firms in the market, offering \( a_1 = a_2 = \frac{4}{3} \)
applications, and there is a monopoly outcome with one firm offering \( a_1 = 3 \), and the
other offering \( a_2 = 0 \) applications (or vice versa). This can be checked directly from
the profit functions.

In this simple model it is almost inevitable that tipping to a single provider is
efficient. That is because the single provider saves duplicated costs of applications,
and, since the applications are marketed to all consumers, it provides more of them.
The duopoly has the benefit of lower prices, but in this model there is no extensive
margin (all consumers are in the market), so the lower prices do not avoid
deadweight loss due to exclusion.
References and Further Reading


*Sega v. Accolade.* 977 F.2d 1510 (9th Cir. 1992).
