Sloan School of Management
Massachusetts Institute of Technology

15.013 – Industrial Economics for Strategic Decisions
Professor Robert S. Pindyck

Lecture Notes
on
Network Externalities

(Revised: August 2012)

These lecture notes will cover some of the more analytical parts of our discussion of markets with network externalities. We will focus largely on situations in which competing firms have different "networks" or "standards." Examples of this include VHS versus Beta in video cassette recorders, the more recent battle between HD-DVD and Blu-Ray in high-definition DVD players, analog versus digital protocols for cellular phone systems, Nintendo versus Sony versus Microsoft in home video games, and Wintel versus Mac in computers.[1] We will consider how firms in markets such as these should think about their pricing, advertising, and product design decisions, and how such markets are likely to evolve.

Although we will focus largely on markets with competing "systems," you should be aware that network externalities exist even when there is a single system. Examples include communications networks (e.g., the telephone system), fax machines, e-mail, and vaccinations for influenza, measles, and other communicable diseases. In each of these cases, the marginal social benefit of one more person joining the "network" is greater than the marginal private benefit. Hence, the equilibrium network size can be less than the socially optimal size, so that the competitive market will not be efficient.

You should also be aware that in markets such as these, there can be *multiple equilibria*, even if all consumers have rational expectations and can fully observe the behavior of other consumers. To see this, consider two possible equilibria for a product such as a fax machine.

[1]For a history of network battles in home video game consoles, starting in 1970, go to www.thegameconsole.com.

1. First, suppose each consumer thinks that no other consumer will buy a fax machine. In this case, no consumer will want to buy a fax machine, so that each consumer is *correct* in his or her expectation regarding the purchase decisions of other consumers. We therefore have a self-fulfilled expectation—each consumer has made a *rational* decision not to buy a machine.

2. Alternatively, suppose each consumer believes that many other consumers will buy fax machines. Then, each consumer will want to buy a fax machine, and indeed many consumers will buy one. We now have a very different self-fulfilled expectation—many consumers have made a rational decision to buy a fax machine, correctly expecting that other consumers will likewise buy them.

An example of a market where we saw the first equilibrium is digital audio tape (DAT). Many consumers "waited" for other consumers to buy digital audio tape players, and eventually concluded (correctly) that very few consumers will ultimately buy them. Hence the product failed to "catch on." An example of the second equilibrium is compact disks. Even though initially the amount of audio "software" available on compact disk was limited, consumers expected other consumers to buy the players, and thus expected that more titles would ultimately be available. This was a self-fulfilling expectation, and compact disk players quickly saturated the market. More recently, the same thing has happened with DVD players (the hardware) and DVDs (the "software").

1 Competing for a Prize

Suppose that two firms are developing competing "systems." It may be the case that ultimately, consumers will all flock to only one of these two systems. Alternatively, the market might ultimately be shared by the two systems, although one might achieve dominance over the other. In either case, these two firms are essentially in a competition for a prize. They can increase their chances of winning the prize by spending more money on product enhancements, on advertising, or (equivalently) on price reductions. At issue is how much money each firm should spend, knowing that its competitor is facing a similar decision.

You have been looking at this problem in the context of the "DOS/MAC" exercise, in which you experiment with a model of dynamic competition with network externalities. However, at this point we can get some insight into what is going on by considering a simpler static problem.[2]

Suppose that two firms compete by choosing how much to spend (on advertising, price reductions, or whatever). Firm 1 spends x, and Firm 2 spends y. Whichever firm "wins" gets W, and the one that loses gets the smaller amount L. Firm 1's chance of winning is $x^a/(x^a + y^a)$, with $a \geq 0$. Likewise, Firm 2's chance of winning is given by $y^a/(x^a + y^a)$. Note that if $a = 0$, the probability that each firm will win is 1/2, independent of how much the firms spend. However, as a gets larger, small differences between x and y result in larger and larger differences in the probability of winning. Thus a large value of a implies that extra effort can have a large expected payoff.

To determine the optimal amount of money that each firm should spend in its attempt to win, we find the Nash equilibrium for this game. Firm 1 takes y as given, and chooses x to maximize its expected net gain:

$$\max \mathcal{E}(\pi) = \frac{x^a}{x^a + y^a} W + \left[1 - \frac{x^a}{x^a + y^a}\right] L - x \tag{1}$$

Taking x as given, Firm 2 chooses y in the same way. Hence Firm 1's reaction function is given by:

$$\frac{a x^{a-1} y^a}{(x^a + y^a)^2}(W - L) - 1 = 0 \tag{2}$$

and Firm 2's reaction function is given by:

$$\frac{a x^a y^{a-1}}{(x^a + y^a)^2}(W - L) - 1 = 0 \tag{3}$$

Combining these two reaction functions, we get the following symmetric Nash equilibrium:

$$x^* = y^* = \frac{a}{4}(W - L) \tag{4}$$

Thus, the total net profit for the two firms is $W + L - (a/2)(W - L)$.

---

[2]This example is from S. M. Besen and J. Farrell, "Choosing How to Compete: Strategies and Tactics in Standardization," *Journal of Economic Perspectives*, Spring 1994.

Note that the more skewed are the returns, i.e., the greater is the difference between W and L, the smaller will be the total joint profit. The reason is that the more skewed the returns, the greater is the incentive for each firm to dissipate potential profits by spending money to increase its probability of winning. Likewise, the larger is a, the smaller will be the total joint profit. The reason is that with a large value of a, even a small difference between x and y can lead to a big difference in the probability of winning, so that once again each firm dissipates profits by spending more money on trying to win.

Suppose, for example, that W=$1000, L=0, and a=1. Then each firm will spend $250 trying to win, and total net profits will be $500. If, on the other hand, a=2, each firm will spend $500 trying to win, and total profits will be 0. Finally, if a=3, each firm will spend $750, and total profits will be -$500, i.e., one firm will win and have a net gain of 1000-750=$250, and the less fortunate firm will simply lose $750.

Is it rational to compete if a>2 so that total profit is negative? Knowing that your expected profit is negative, you might simply decide it is better not to compete. Or, you might decide that your competitor has reached this very same conclusion and will not compete - in which case, you *should* compete, because you will be sure to win. In other words, you might find yourself in a War of Attrition.

We will see that as a general matter, in markets with skewed returns, where a small advantage by one firm can "tip" the scales, competition can be very intense. Both firms are essentially competing to win a future monopoly position. Both firms might even have an expected net payoff that is negative, but will compete anyway in the hope of winning the War of Attrition.

2 Dynamic Models of Pricing with Network Externalities

We now turn to dynamic models in which firms have competing "networks," and there is a process of saturation. As we will see when we discuss pharmaceutical markets, models of this sort can be estimated using regression methods, and thereby used to measure the extent of network externalities. These models can also be used to simulate the effects of different pricing strategies.

2.1 Diffusion, Saturation, and the S-Shaped Curve.

We will be utilizing diffusion models that often appear in marketing applications, so it might be helpful to begin by reviewing these models and how they work in a monopoly context. For those of you who never encountered this in your marketing courses, here is a brief overview of a diffusion/saturation process for a new, semi-durable products (such as an Apple iPod). Suppose that X_t is the *stock* (not the level of sales) of iPods in circulation at time t, and X^* is the *saturation level*, i.e., the maximum stock that we think could be in circulation after every potential iPod customer has bought one. For the moment, lets assume that X^* is a fixed number that does not change over time.

The *actual* stock of iPods in circulation might follow an equation such as the following:

$$X_t = X_{t-1} + \alpha X_{t-1}(\ln X^* - \ln X_{t-1}) \tag{5}$$

This equation says that the stock of iPods this period (e.g., this year or month) will equal the stock last period plus the growth in the stock. The growth in the stock depends on two factors: (1) last periods stock, because the greater is the number of iPods in circulation, the greater will be the consumer awareness of the product; (2) the potential unsatisfied demand for iPods, as represented by the difference between the log of the saturation level and the log of the stock in circulation. The parameter α determines how rapidly this process of diffusion and saturation occurs. The solution of this equation is the famous S-shaped curve for the life cycle of a new product. (See Figure 1 below for an example.)

What about sales? Sales each period are equal to the increase in the stock plus replacement sales (for lost, broken, or worn-out iPods). Replacement sales are equal to a depreciation rate (δ) times the previous stock. Thus the equation for sales is:

$$S_t = X_t - X_{t-1} + \delta X_{t-1} \tag{6}$$

Note that sales will start out slowly, then increase as the stock increases, and later decrease and level out as the market saturates. Once the market has saturated, the only sales will be replacement sales. (See Figure 2 below.)

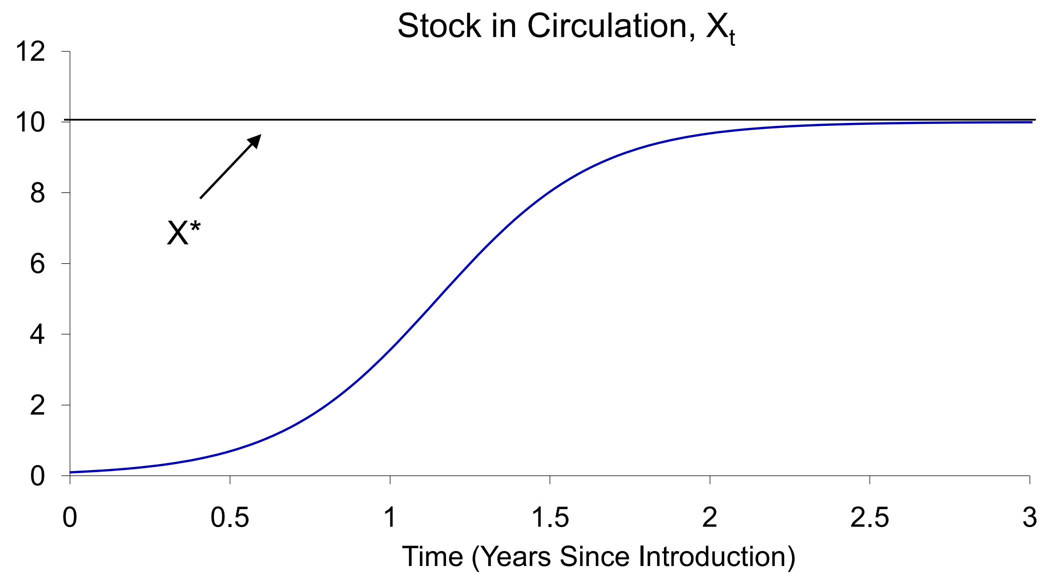

Figure 1: Stock in Circulation (S-Shaped Curve)

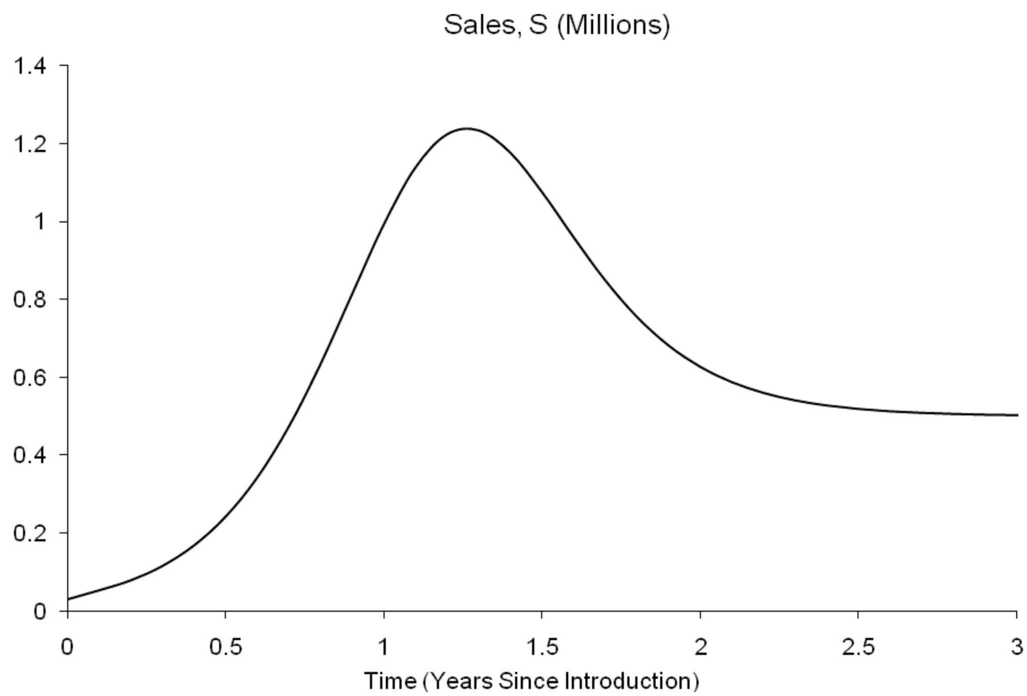

Figure 2: Rate of Sales (Units sold per period)

There is no reason, however, for the saturation level X^* to remain fixed. We would expect the saturation level to depend on GDP (as the economy grows, the potential market for iPods will grow) and on price (if the price of iPods falls, the potential market will grow). Thus we might make the saturation level a function of GDP, price, and perhaps other variables:

$$\ln X_t^* = a_0 - a_1 \ln P_t + a_2 \ln \mathrm{GDP}_t \tag{7}$$

Here, the parameters a_1 and a_2 are long-run price and income elasticities. Now the saturation level will change over time as price and GDP change.

2.2 Competing Firms.

Let $x_1(t)$ be the number of consumers who are using the "network" of Firm 1, and $x_2(t)$ be the number that are using the "network" of Firm 2. (By "network" we mean a product with network externalities, such as a Wintel versus Mac computer, one brand of anticholesterol drug versus another, or a Nintendo video game system versus a Sony system.) Let x^* be the saturation level, which we will assume, for simplicity, is fixed. Then, if market saturation follows a logistic curve, the dynamics of x_1 and x_2 are given by the following two equations:

$$\Delta x_1(t) = \alpha_1 x_1(t-1)[x^* - x_1(t-1) - x_2(t-1)] \tag{8}$$

$$\Delta x_2(t) = \alpha_2 x_2(t-1)[x^* - x_1(t-1) - x_2(t-1)] \tag{9}$$

Here, $\alpha_1 = \alpha_1(P_1, P_2; A_1, A_2)$ and $\alpha_2 = \alpha_2(P_1, P_2; A_1, A_2)$, where P_1 and P_2 are the prices of each product, and A_1 and A_2 are the firms' advertising levels. In other words, the *rate of saturation* for each competing product is a function of relative prices and relative advertising levels.

For example, α_1 and α_2 might be given by:

$$\alpha_1 = a_1(P_1/P_2)^{-\beta}(A_1/A_2)^{\gamma} \tag{10}$$

$$\alpha_2 = a_2(P_1/P_2)^{\beta}(A_1/A_2)^{-\gamma} \tag{11}$$

Note that if $P_1 > P_2$, this gives Firm 2 an advantage, so that its rate of saturation will be faster, and it will gain market share at the expense of Firm 1.

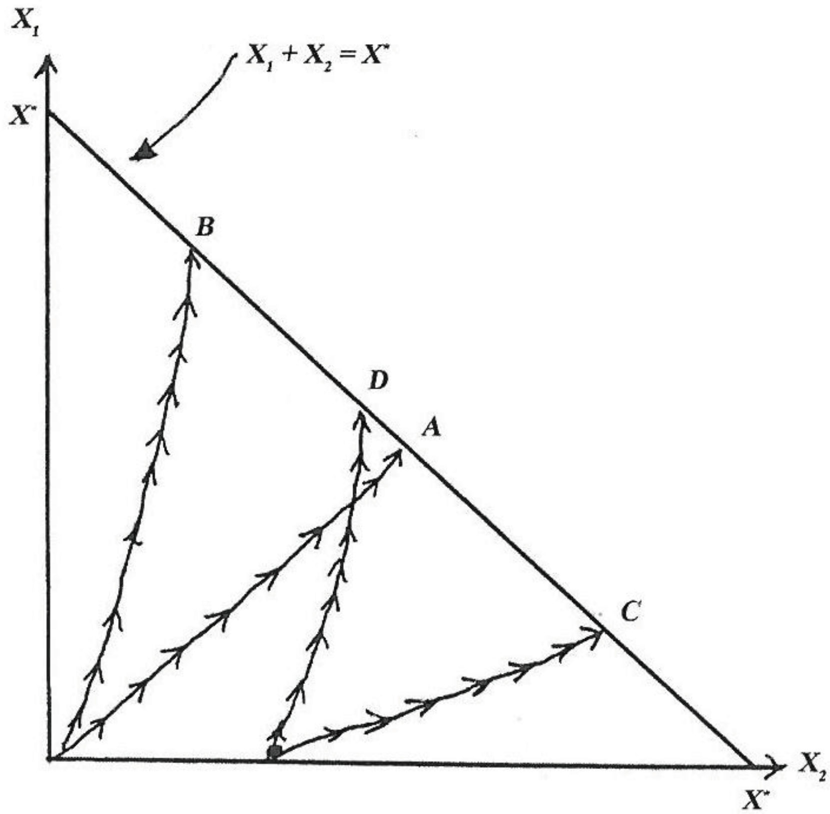

Figure 3: Possible Trajectories for Market Shares

Figure 3 shows some possible outcomes of this competition. For the two trajectories ending at points A and B, both firms begin with equal numbers of consumers that have adopted their products, although this number is small. In Trajectory A, prices and advertising levels are the same for the two firms, so that both firms increase sales proportionately, and by the time the market saturates, they have equal shares. In Trajectory B, both firms once again begin with equal shares, but $\alpha_1 > \alpha_2$, perhaps because Firm 1 is charging a lower price and doing more advertising. Hence, Firm 1's sales increase faster than those of Firm 2, and by the time the market has saturated, Firm 1 has a much larger share.

In Trajectory C, both firms set the same prices and advertising levels, so that $\alpha_1 = \alpha_2$. However, Firm 2 starts out with a much larger share than Firm 1. Observe that this advantage will translate into an advantage at the time of saturation; Firm 2 will have a much larger share of the saturated market. In Trajectory D, Firm 2 again begins way ahead of Firm 1. But now, $\alpha_1 > \alpha_2$, perhaps because Firm 1 is charging a much lower price or doing much more advertising. In this case, Firm 1 picks up a considerable share of the market, so that by the time of saturation it has more than a 50-percent share.

To summarize, the four trajectories are as follows:

$$A: \quad x_1(0) = x_2(0); \quad \alpha_1 = \alpha_2$$

$$B: \quad x_1(0) = x_2(0); \quad \alpha_1 > \alpha_2$$

$$C: \quad x_1(0) < x_2(0); \quad \alpha_1 = \alpha_2$$

$$D: \quad x_1(0) < x_2(0); \quad \alpha_1 > \alpha_2$$

2.3 How to Compete?

Suppose that you are Firm 1, and you are competing against Firm 2, with both of you starting at small but equal shares. Should you price aggressively in order to move along Trajectory B, or should you generally match Firm 2's price and advertising level, in the hope of moving along Trajectory A? This is a difficult question to answer. The problem is similar to the "Competing for a Prize" example that we looked at earlier, but more complicated because of the dynamic structure of the saturation process. First, even if you could somehow predict the prices that Firm 2 will charge in the future, your optimal decision would depend on how your prices affect the NPV of your flow of profits. On top of this, you cannot know what Firm 2's prices will be. Both firms are locked in a dynamic game, where they are continually competing with each other by adjusting prices and advertising levels.

This game can have many different equilibria. As with the repeated Prisoner's Dilemma, one possibility is that the two firms will "agree" not to compete aggressively, perhaps by using some kind of tit-for-tat strategy. But another possibility is very aggressive competition in which the two firms drive each other into the ground trying to win a large share of the market. This second possibility is more likely if the rates of saturation, α_1 and α_2, are highly sensitive to relative prices; in that case each firm could suffer a large permanent loss of market share if it lets down its guard and allows itself to be undercut by its competitor.

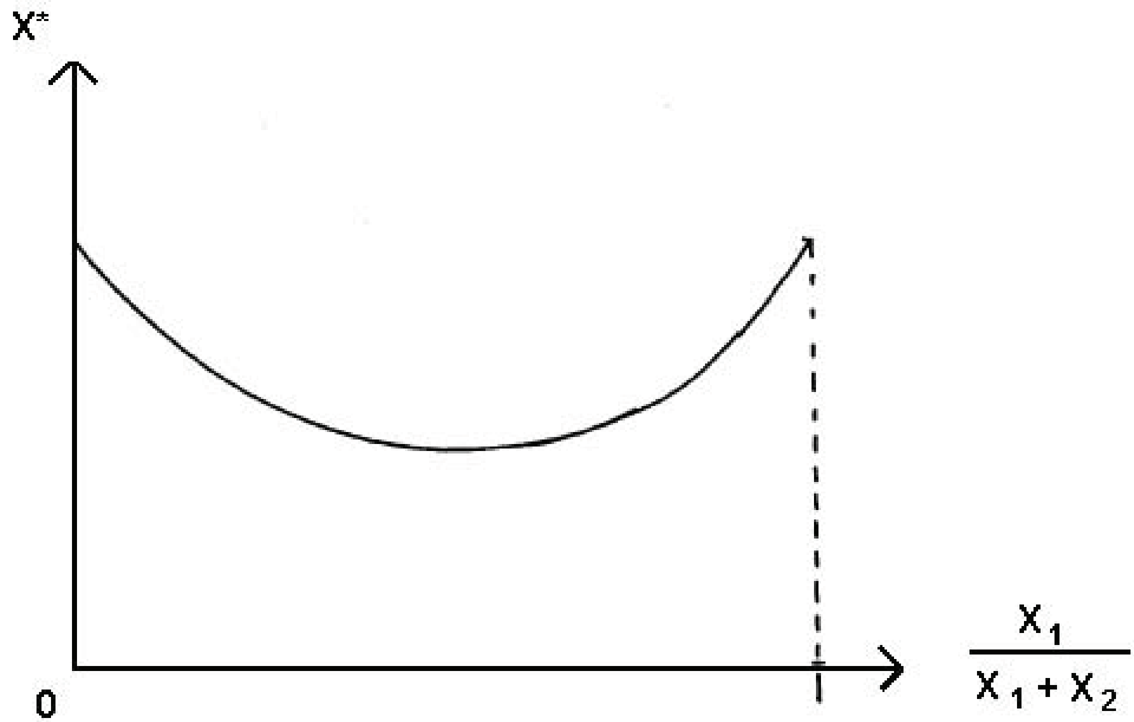

Figure 4: X^* Depends on Market Split. Note X^* is largest if one firm has entire market.

Note that in Figure 3, the saturation level x^* is independent of the market shares for each firm. With network externalities, x^* might, in fact, be larger if one firm dominates the market. An example of this is video cassette recorders. Imagine a world in which the Beta and VHS formats each had a 50-percent share of the market. That would increase the costs of prerecorded video cassette rentals (stores would have to stock movies in two formats), make it more difficult for consumers to exchange tapes, etc. The value of a VCR for the typical consumer would thus be lower, and the saturation level would likewise be lower.

Figures 4 and 5 illustrate this. Figure 4 shows how x^* might depend on the market split. In that figure, x^* is greatest when *either* firm has the entire market, i.e., when $x_1/(x_1 + x_2)$ is either 0 or 1. Figure 5 shows how Figure 3 would then be modified.

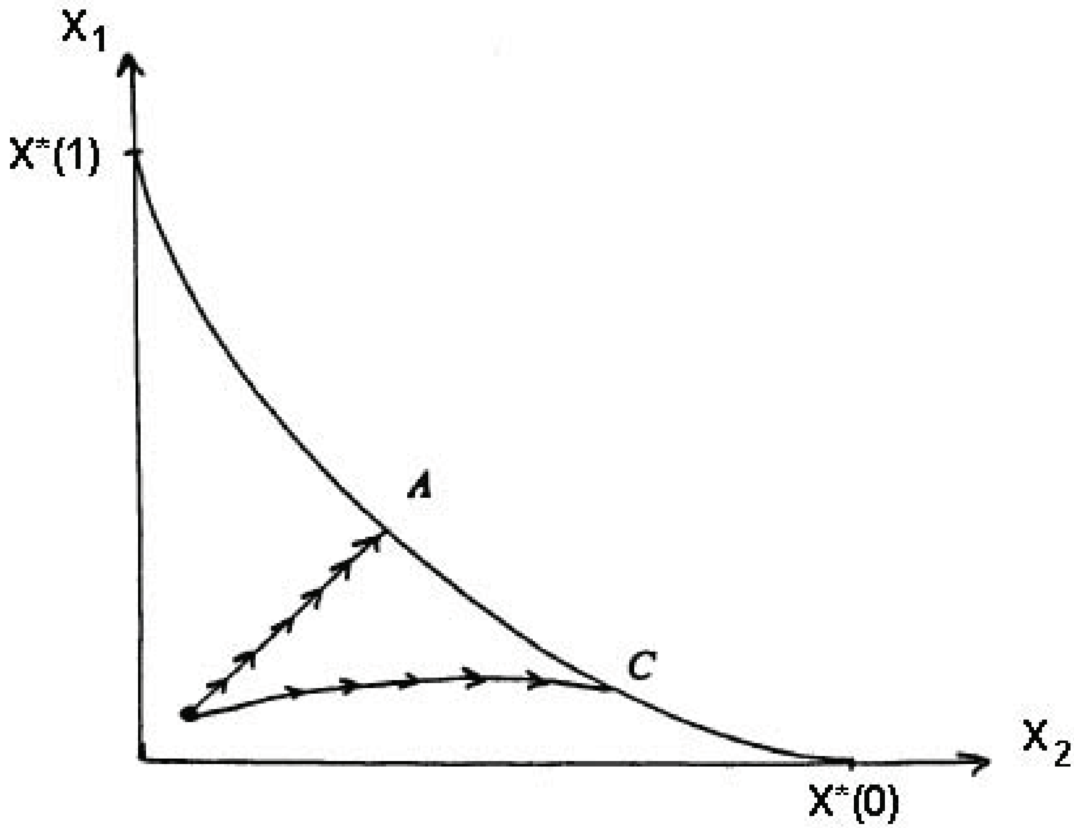

Figure 5: Trajectories for Market Share. Note total sales (X^*) is larger at Point C (where Firm 2 has most of the market) than at Point A (where the market is evenly spilt).

2.4 The Laggard's Dilemma

Let's return to Figure 3 for a moment. Suppose once again that you are Firm 1, but you currently have a very small user base, while Firm 2 has a large user base. In other words, you are are the laggard — you are starting from the beginning of Trajectory C or D, with a small market share. Now, what should you do? Should you fight hard, reducing your price and increasing your advertising dramatically, even if that means losing a great deal of money for a while? After all, maybe you can succeed in gaining market share (so that you follow Trajectory D). On the other hand, perhaps you should you effectively give up and cede the market to Firm 2. This way you can maintain a high price, and although your market share will remain small, you can "milk" the product and earn at least some profits (and follow

Trajectory C),

If you thought you could drop your price (or increase your level of advertising) without Firm 1 responding in kind, the first strategy might be a good one. If the network externalities are not too strong, your low-price strategy might allow you to gain market share, and even become the market leader after some time. But do you really think that Firm 1 would not respond to your drop in price? After all, Firm 1 has a lot to lose by giving up its market dominance. Furthermore, if the network externalities are strong, Firm 1 would not even have to match your price — it could reduce its price somewhat, and rely on the strength of the network externality to avoid losing share.

Once again, you face a complicated dynamic game that can have many outcomes, and your choice of strategy depends on the kind of response you expect from your competitor. However, it turns out that deciding what to do when you are far behind is typically much easier than deciding what to do if you are close to even with your competitor. Here is the solution to the "Laggard's Dilemma:" Unless you believe that your competitor will not respond in kind to a price cut or other attempt to gain share, *you should milk your product as best you can and cede the market to your competitor.* Why? Because with moderate to strong network externalities, you should expect to lose much more money than your competitor in a fight for market share. Thus, with a rational competitor, you face an uphill battle that you will almost surely lose, and at great expense.

In Exercise 9, you will have a chance to explore the "Laggard's Dilemma," and see for yourself what kinds of strategies work and don't work when you are coming from behind. You will do this in the context of a "DOS-MAC" simulation, in which you will face Apple's situation in the mid-1980's.

3 Connectivity and Compatibility

In markets where network externalities are strong, connectivity and/or compatibility can be important factors in determining current or potential future market power. Furthermore, there are times when a firm with a large market share can expand its share by choosing to *selectively degrade* its connectivity with other firms' networks, or its compatibility with other firms' products.

3.1 Connectivity

Consider a set of firms that compete with each other in providing some service over a network. An example would be wireless (cellular) telephone service. AT&T, Sprint, and Verizon all provide wireless telephone service, and each operates its own network. The networks, however, are completely connected: a consumer who happened to be an AT&T subscriber, for example, could easily make a call to a friend who happened to be a Verizon subscriber (and, unless told so, would not even know that the friend was a Verizon subscriber). This complete connectivity is extremely important from the point of view of consumer welfare: imagine a world in which a Verizon subscriber could only call another Verizon subscriber, an AT&T subscriber could only call another AT&T subscriber, etc. Furthermore, this complete connectivity helps to ensure that there will be strong competition among wireless providers. With complete connectivity, it is difficult or impossible for any one provider to dominate the market, and thus providers are forced to compete along price and quality dimensions.

Complete connectivity cannot always be taken for granted, however. In principle, a firm with a large market share could decide to "selectively degrade" the connectivity of its network with those of its competitors. Such a strategy would be detrimental to consumers, but could increase the market share, market power, and profitability of the dominant firm.

To see how this could occur, let's examine the Internet backbone market. (You might want to review our more detailed discussion of this market in R. Pindyck, "Notes on Internet Economics and Market Structure.") There are currently some five to ten major Internet background providers (IBPs), including Cogent Communications, Sprint, AT&T WorldNet, and Level 3, which recently acquired Global Crossing, among others. (There are also quite a few smaller, and largely regional, backbone providers.) These companies transmit data over long distances and over large regions of the world, typically using long-haul fiber optic cables.

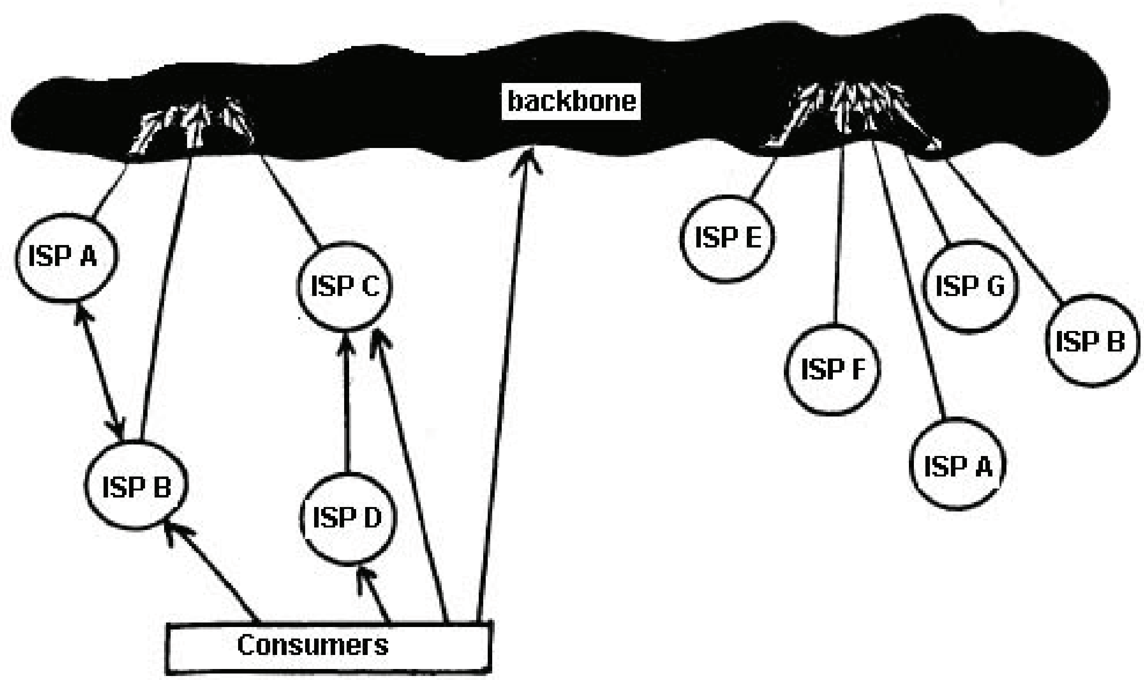

Figure 6: The Internet Backbone Market

Figure 6 shows the general structure of Internet communications.[3] Most consumers access the Internet through an Internet Service Provider (ISP), sometimes through an ordinary telephone line, but much more commonly through a cable or a DSL line. The connections among ISPs then occur (most of the time) via the Internet backbone. For example, your ISP might obtain backbone service from Level 3, while your friend's ISP obtains backbone service from Sprint.

Some institutions obtain Internet service directly through a backbone provider. For example, Barnes&Noble has connected directly to Cable & Wireless, while MIT is connected directly to Level 3 (which now owns the IBP assets of GTE).[4] But now a problem arises. Suppose that you want to buy a book online from Barnes&Noble and you do so by accessing their Web page (www.bn.com) from a computer at MIT. In this case, information must go

[3]This figure is from J. Crémer, P. Rey, and J. Tirole, "Connectivity in the Commercial Internet," unpublished, April 1999. The discussion that follows is based in part on that paper.

[4]Cable & Wireless is being acquired by Vodafone.

back and forth between the Cable & Wireless and the Level 3 networks. Clearly, some kind of connectivity is necessary to make that happen.

So far there has been a very high degree of connectivity across Internet background providers. The fact that you can easily order things from Barnes&Noble — and myriad other websites — using an MIT computer is evidence of this. This connectivity occurs through "peering" agreements. Essentially, there has been an informal agreement among IBPs to "peer" with each other, i.e., to transmit each other's messages. The question, however, is whether this high level of connectivity will persist. This question came up during the attempted merger of Sprint with MCI/Worldcom several years ago. The answer was: maybe not. This was one of the most important reasons that the proposed merger was blocked.

To understand the problem, note that Internet backbone providers have very high sunk costs (for installing fiber optic cables, switching circuits, etc.) and very low marginal costs, and sell a homogeneous product (after all, a bit is a bit). This creates intense price competition, to the considerable benefit of consumers. However, it has serious implications for the economic profitability of the backbone providers.

Suppose that over time, one of these providers somehow gains a large market share relative to the others. (This clearly would have happened, for example, had the Sprint and MCI/Worldcom merger gone through.) Suppose that backbone provider A has a 40-percent market share, and the other three providers each have a 20-percent market share. What would happen if provider A decided to "selectively degrade" its connectivity with the other IBPs? (There are a variety of ways that this could be done: for example, by allowing the messages from other IBPs to queue up before sending them on to their final destinations, thereby creating delays.)

In such a situation, there are two possible outcomes. First, suppose that backbone providers B, C, and D have excellent connectivity among themselves. In that case, most ISPs and large companies would have an incentive to connect to one of them, because that way they would have 60-percent coverage, and a lower chance of a connectivity problem than if they use provider A. As a result, the market share of provider A will shrink, giving companies an even greater incentive to use providers B, C, or D. Unless it improves its connectivity with B, C, and D, provider A might eventually go out of business.

On the other hand, suppose that providers B, C, and D have good, but not perfect, connectivity among themselves. (This is the more realistic scenario.) In that case, most ISPs and large companies would have an incentive to connect directly to backbone provider A, and avoid the other three firms. Why? Because with provider A, there is a lower likelihood (on average) of having to connect with another provider, and thus a higher likelihood that messages will be transmitted and received quickly. But this, in turn, would mean that the market share of provider A will grow, perhaps from 40 percent to 60 percent. Then, there will be an even greater incentive for ISPs and large companies to connect to provider A, so that its share grows even more. Eventually, this could lead to "market tipping," where provider A becomes a monopolist or near-monopolist. In such a situation, consumers would have no problems of connectivity, but would face higher prices for the use of the Internet.

3.2 Compatibility

Analogous issues arise in the case of product *compatibility*. This is easiest to see in the context of applications software (although it occurs in other situations as well). Suppose that there are four competing word processing programs, and suppose that these programs are *completely compatible*. By this I mean that a document file created in Program A could be perfectly and effortlessly read by Programs B, C, and D. In other words, if you and I were co-authoring a document, and you were using Program A and I were using Program B, we could send the file back and forth and never be concerned that we were using different programs. (Perhaps I chose Program A because of its price and/or characteristics, and you chose Program B for similar reasons.)

There may or may not be a network externality that could lead, at least to some degree, to tipping towards one program or another. For example, suppose that the four programs have different command structures, and that Program A has a 40-percent market share while Programs B, C, and D each have a 20-percent share. Other things equal, some consumers might prefer Program A because its larger market share would mean that it would be easier to find people to help them use it.

With *imperfect compatibility*, however, the tendency for market tipping becomes much stronger. Suppose, for example, that Program A has a 40-percent market share, Programs B, C, and D have 20-percent market shares each, and the programs are *completely incompatible* with each other. (In other words, a document created with Program A cannot be read by Programs B, C, or D, and vice versa.) In this case, the benefits to using Program A become much greater, because of the greater likelihood of being able to exchange files with other people. Thus more consumers will migrate to Program A, giving it an even larger market share. This will increase the benefits of using Program A even more, its market share will increase further, and so on.

Currently, Microsoft has a roughly 90-percent share of the market for word-processing software. Microsoft Word is largely compatible with WordPerfect and other word-processing programs, but not completely compatible. (If a WordPerfect file is opened in Microsoft Word, equations, symbols, footnotes, and highly formatted text will often transfer incompletely, incorrectly, or not at all.)

Suppose you were advising Microsoft regarding plans for the compatibility of the next version of Word with other word processing packages. Should Microsoft strive to increase the compatibility of Word with other packages, or instead selectively degrade the compatibility?

4 Network Externalities in Markets for Pharmaceuticals

In pharmaceutical markets, network externalities can be of two types. First, there is a network externality associated with a *therapeutic category*. Examples of therapeutic categories include H_2-antagonist antiulcer drugs, SSRI antidepressants, and statin-type anticholesterol drugs. For any one of these therapeutic categories, both doctors and patients will be more willing to adopt a drug in the category if a large number of other patients have taken or are taking a drug in the category. In other words, doctors are more willing to prescribe, and patients to take, a type of drug if that type of drug has been "accepted," and "acceptance" is best measured by the number of other people that have taken or are taking a drug in that therapeutic category.

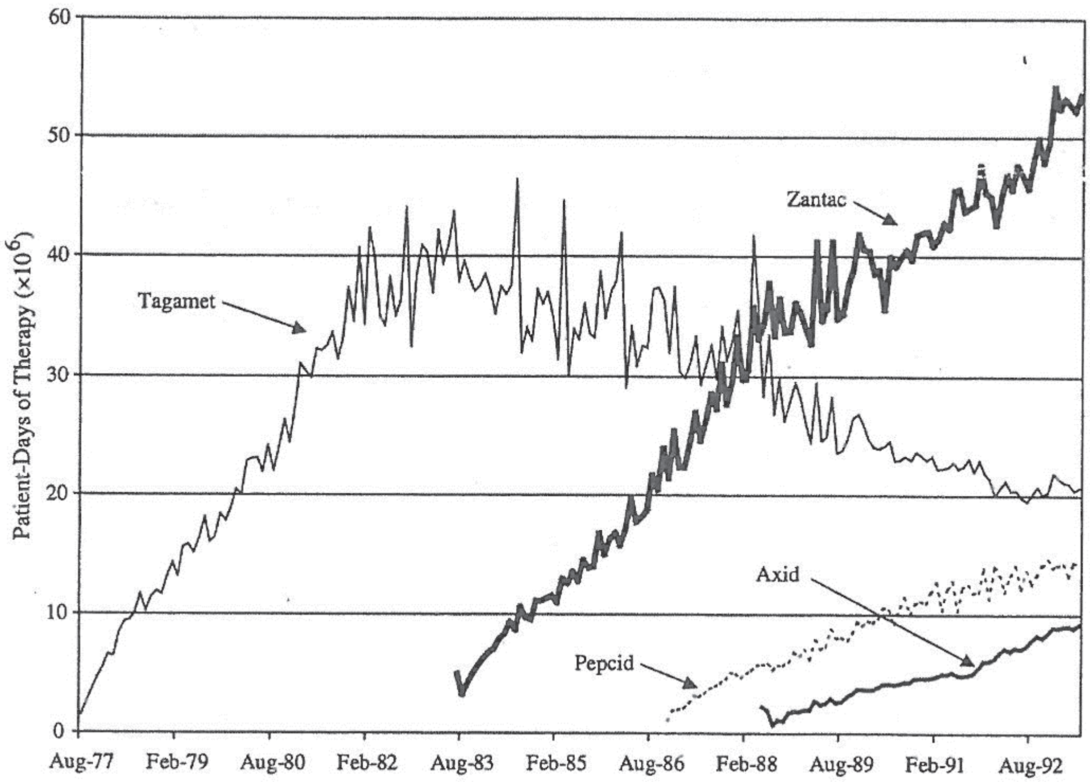

Figure 7: Monthly Sales for H_2-Antagonist Drugs

Pharmaceutical companies are also concerned with the possibility of a second network externality, which is associated with the *brand* of drug within the therapeutic category. If this network externality is important, it means that doctors and patients would be more willing, for example, to use Zantac (as opposed to one of the very similar drugs, Tagamet, Axid or Pepcid) the greater the market share of Zantac. This is an important issue for pharmaceutical companies, because if the brand-specific network externality is large, it means that the owner of a dominant brand will have market power, and can afford to raise prices above those for brands with smaller shares. It would also mean that the reward for being first in the market is large, so that it is worthwhile for a company to spend a good deal of money to accelerate the development of a new drug.

Let's look in more detail at the H_2-antagonist antiulcer drugs. The pioneering drug in this therapeutic category is Tagamet, which was released in 1977, and revolutionized

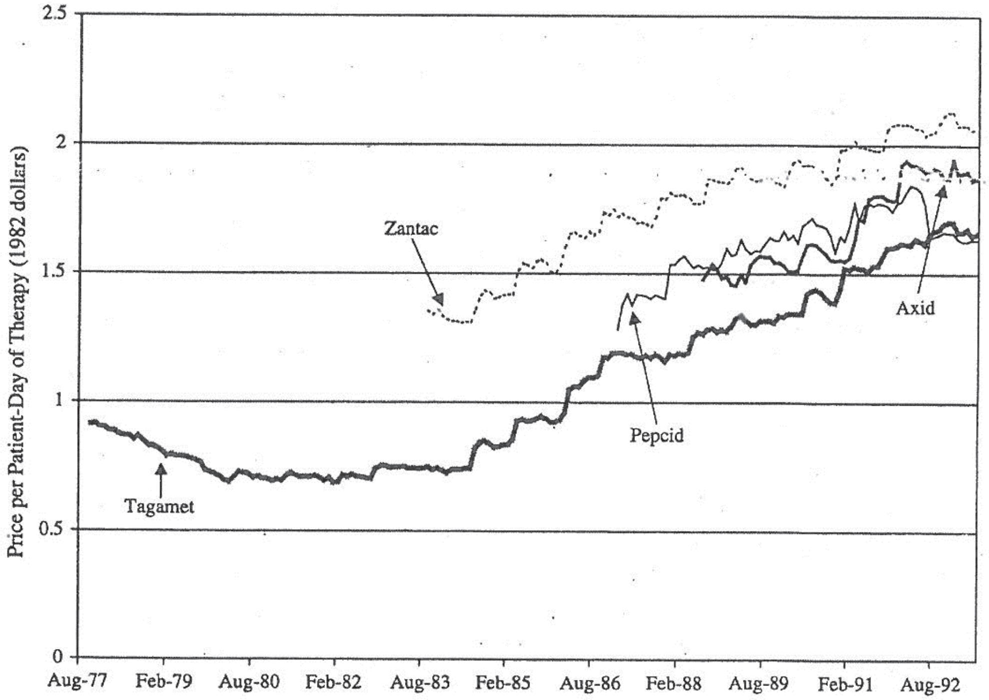

Figure 8: Real Prices of H_2-Antagonist Drugs

the treatment of stomach ulcers. Three similar drugs followed: Zantac, Axid, and Pepcid. Figures 7 and 8 show sales and (real) prices for the four drugs in this category. Note that Zantac, which came into the market about six years after Tagamet, had a rapidly rising share, despite the fact that it was priced well above Tagamet. In part, this is due to the somewhat superior attributes of Zantac relative to Tagamet, e.g., fewer doses per day and fewer interactions with other drugs. Nonetheless, the fact that Zantac overtook Tagamet is evidence that brand-specific network externalities in the market for H_2-antagonist antiulcer drugs, although present and statistically significant, are weak.[5] If this were not the case, we would have expected to see Tagamet retain its market share for much longer.

[5]The statistical significance of a brand-specific network externality is based on a study by Ernst R. Berndt, Robert S. Pindyck, and Pierre Azoulay, "Consumption Externalities and Diffusion in Pharmaceutical Markets: Antiulcer Drugs," *Journal of Industrial Economics*, June 2003.

5 A Standards Battle: High-Definition DVDs

We discussed at some length the standards battle that emerged nearly three decades ago when videocassette recorders were first introduced. That battle between the VHS and Beta formats was eventually won by VHS (even though Beta was a better technology). When the first-generation DVD players were developed and introduced, a repetition of that standards battle was averted — Sony and Philips gave up on their own technology standard, and instead adopted the Toshiba standard in exchange for a share of royalties. The result was that consumers faced no confusion over DVD standards, and DVD hardware and software rapidly penetrated the market.

But in 2005 a standards war emerged once again, this time over the new generation of high-definition DVDs. Toshiba developed and continued to promote its HD-DVD standard, and Sony developed and continued to promote its Blu-ray standard. (The two standards are incompatible; HD-DVD disks will not play on a Blu-ray machine, and Blu-ray disks will not play on a HD-DVD machine.) The battle continued for about two years, with some movie studios lining up behind one format and other studios lining up behind the second format. The result was that very few machines (HD-DVD or Blu-Ray) were sold, and very few movies were made available in either format. Given the lack of "software" and the uncertainty over which format would eventually "win," consumers decided it was best to wait rather than buy a machine. For a while, it looked like this format battle would prevent high-definition DVDs from ever taking off.

The battle ended when most movie studios lined up behind Blu-ray and decided to stop issuing movies in the HD-DVD format. On February 19, 2008, Toshiba announced that it would stop development of the HD-DVD players, and conceded the battle to the Blu-ray format. As Table 1 shows, sales and rentals of high-definition DVD movies have now started to grow. But note also that in the past few years, *total* sales and rentals of DVDs has declined, in large part because of the increasing availability of video on demand (via cable and satellite television), streaming video, and the ability to record movies and other video from television to a DVR box.

Table 1: **DVD Sales and Rentals ($ Billions)**

Year	VHS	DVD	High-def DVD	Total
1999	$12.5	$0.8	$0.00	$13.3
2000	$11.8	$2.5	$0.00	$14.3
2001	$11.1	$6.8	$0.00	$17.9
2002	$9.1	$11.6	$0.00	$20.7
2003	$6.1	$16.1	$0.00	$22.2
2004	$3.7	$21.2	$0.00	$24.9
2005	$1.6	$22.8	$0.00	$24.4
2006	$0.4	$24.1	$0.00	$24.5
2007	$0.1	$23.3	$0.27	$23.7
2008	$0.0	$21.6	$0.75	$22.4
2009	$0.0	$18.1	$2.00	$20.1
2010	$0.0	$15.8	$2.30	$18.1

6 Network Markets: Credit Cards

There are very few markets in which network externalities play a stronger role than the market for credit and charge cards. The nature of the (positive) network externality should be easy to see: You might carry and often use a Visa card because lots of merchants accept the card, making it easy for you to make purchases. Likewise, lots of merchants accept Visa cards because such a large number of consumers have them and use them.[6]

This externality creates a "chicken-and-egg" problem that makes starting a new card network very difficult. Consumers will obtain and use the card of a new network only if lots of merchants accept the card, but lots of merchants will accept the card only if many consumers have them. So to start a new network, we must somehow convince *both* consumers and merchants to obtain, use, and accept the cards of that network. This is a much harder job than it would be if all we had to do was convince *either* consumers *or* merchants to adopt the card. Note the difference, for example, between card networks and computer

[6]These lecture notes provide a bare-bones introduction to card networks and how they work. For an excellent and thorough treatment, see David Evans and Richard Schmalensee, *Paying with Plastic: The Digital Revolution in Buying and Borrowing*, 2nd Ed., MIT Press, 2004.

productivity software. All Microsoft had to do was get many or most computer users to buy and use Office — not an easy job, but once enough people were using Office, the network externality would take over and more and more people would use Office. With a credit card network, we must get two completely different groups to accept and use the card, and each group will do so only if the other group does.

6.1 Charge Cards, Credit Cards, and Debit Cards

There are three major categories of cards that can be used to buy goods and services: charge cards, credit cards, and debit cards. The focus of these notes will be on charge and credit cards, but it is important to understand the differences among all three. All three are sometimes called *payment cards* because they can be used as a form of payment.

A *charge card* provides a transaction service; i.e., it allows consumers and merchants to do business with each other without the need for cash, checks, or other types of payment. An example of a charge card is the American Express card. You can use an Amex card to make purchases, but you are expected to pay your balance due at the end of each month. The very first payment cards, beginning with the Diners Club card introduced in New York in 1950, were all charge cards.

A *credit card* bundles two services together: the transaction service that a charge card provides, along with a credit service. Example, of course, are Visa and MasterCard. About half the people (in the U.S.) who hold one or more credit cards make use of only the transaction service — they pay their balance due in full at the end of each month. The other half of cardholders also use the credit service; they maintain a balance due, on which they pay interest each month.

A *debit card* is something in between a credit card and a charge card. It, too, provides a transaction service, in that you can use the card as a form of payment just as you would a credit or charge card. However, rather than receiving a bill at the end of the month with a balance due, as you would with a credit or charge card, every charge you make with the card is debited directly from your bank account. (Some banks will issue you a debit card that provides some credit, in that the bank will account balance to go negative, and of course charge you interest accordingly.)

There are actually two different kinds of debit card transactions. An *off-line* debit transaction is signature-based, so that you would use the card the same way you would use a credit card. In the U.S., an off-line debit transaction can only be done with a card that is part of the Visa or MasterCard networks. When you make an off-line debit transaction, the merchant gets paid (and your bank account gets debited) in about two or three days.

An *on-line* debit transaction requires a PIN number. Merchants prefer on-line debit because they get paid immediately, and they also pay a much smaller discount. Thus a merchant will often suggest that you enter a PIN number.

The use of debit cards has grown extremely rapidly over the past decade, so that debit card charge volume is now close to credit card volume. This might seem surprising given that debit cards have distinct disadvantages compared to credit cards. The most important disadvantages are (1) your account is debited immediately, eliminating about a month of free credit ("float"), and (2) a mistaken, fraudulent, or otherwise disputed charge is debited from your bank account and stays debited until the dispute is resolved. Then how can we explain the rapid growth of debit cards? First, it is much easier for students and other people without a credit history to obtain a debit card than a credit card. Second, many people fear (sometimes correctly) that with a credit card they will be unable to constrain their spending.

6.2 Open and Closed Systems

There are three important functions that must be carried out for any card systems. (1) Cards must be issued to consumers, and then cardholders must be sent monthly statements, and dealt with when disputes or other problems arise. (2) Merchants must be solicited and signed up to accept the cards, and then dealt with when problems arise. (3) All of the card transactions must be processed, and the processing must be quick. There are two fundamental types of card systems, and they differ based on how these functions are carried out.

American Express and Discover operate *closed systems*. In a closed system, the card company carries out all three functions. Thus American Express solicits and issues cards to consumers, sends its cardholders monthly statements and collects their payments, and deals with lost cards and disputed charges. American Express also signs up ("acquires") merchants to accept the cards, helps them set up the necessary infrastructure (card readers, etc.), makes sure they are paid on time, and deals with any complaints or other issues. Finally, American Express processes all of the transactions involving its cards.

Visa and MasterCard, on the other hand, operate *open systems*. The only fundamental function that they perform is the processing of transactions (often with the help of third parties). (Visa and MasterCard also perform a secondary function of creating brand standards for their cards, advertising their networks, and doing some R&D related to card security.) Then who finds the consumers and merchants? Any bank can join the Visa or MasterCard network (or both) and issue cards. Likewise, any bank can become an "acquirer" and solicit and sign up merchants.

If you carry a Visa or MasterCard credit card, your relationship is with the bank that issued you the card, and *not* with Visa or MasterCard. If you have a dispute over a charge, you take it up with your issuer, not with Visa or MasterCard. And the particular terms for your card (annual fee, possible rebates, the interest rate on unpaid balances) is likewise determined by your issuer. And a merchant that accepts Visa and MasterCard must take up any dispute with its acquiring bank, not with the Visa or MasterCard networks. If you carry an American Express card, on the other hand, your relationship is with American Express, which settles any disputes you might have and sets the terms for your card.

There are advantages and disadvantages to each type of system. (If that were not the case, we wouldn't see two systems.) What are the advantages of an open-loop system such as MasterCard's? What are the advantages of a closed-loop system. Think about these questions, and we will discuss them in class.

Table 2 shows the top U.S. card issuers in 2010, based on annual purchase volume. In terms of the number of cards issues, Chase was number 1 by far, with 130 million cards. American Express had only issued about 50 million cards, but because an Amex cardholder typically spends much more per purchase than a Visa or MasterCard cardholder, it was first in terms of purchase volume.

Table 2: **Top U.S. Card Issuers in 2010**

Rank	Issuer	Purchase Volume ($ Billions)
1	American Express	460.7
2	Chase	333.1
3	Bank of America	237.5
4	Citibank	192.5
5	Capital One	98.3
6	Discover	92.5
7	U.S. Bancorp	73.6
8	Wells Fargo	49.1
9	HSBC	30.6
10	Barclays	29.8

(Source: Nilson Report, February 2011)

6.3 Interchange and the Merchant Discount

What happens when you make a purchase with a charge card or credit card? What is the cost to the merchant (the "merchant discount") when you use a card? The answers depend on whether the card is part of a closed or open system, and which card network it belongs to. Let's start with Visa and MasterCard, which are the largest networks in terms of the numbers of cardholders. Figure 9 illustrates what happens when you make a make a purchase.

Suppose you go to a store and use a Visa card (issued, let's say, by Bank of America) to buy something for $100. You (or the cashier) swipe the card through the card reader. Your card number identifies the issuer of the card, so a "message" is sent to that issuer requesting authorization for the purchase (B). If your account is delinquent, the authorization will be denied, but let's assume that your account is in good standing, in which case the authorization will be approved (C). All of this take two or three seconds. You then leave the store with your purchase.

The store will then request payment from its acquiring bank (D). The acquiring bank will request payment from Visa (E), and Visa will request payment from the issuing bank, in this case Bank of America (F). The issuing bank will immediately pay Visa, but not the

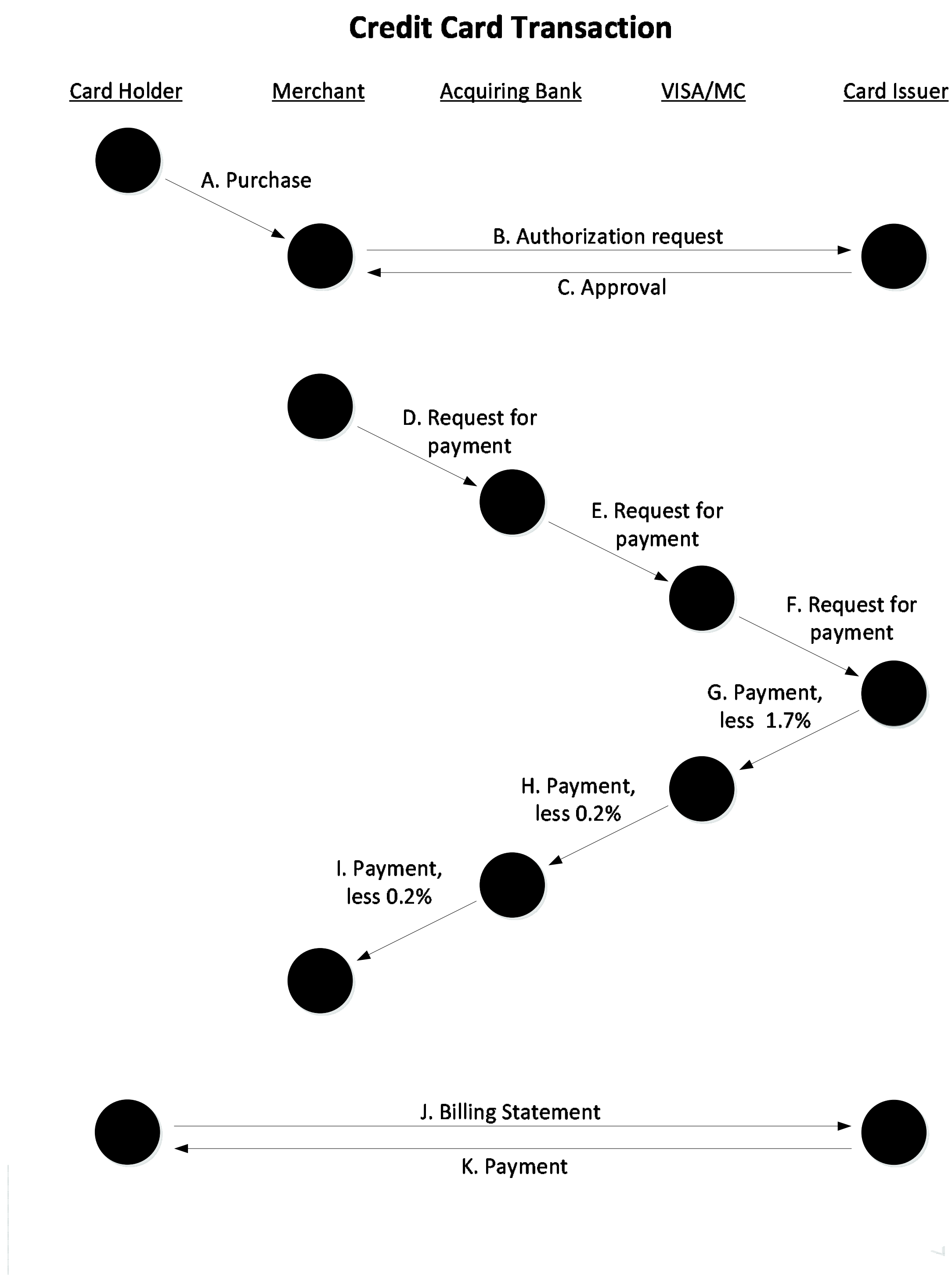

Figure 9: Visa or MasterCard Credit Card Transaction

entire \$100. It will keep a fraction of the \$100 for itself, usually around 1.7% (G). That fraction is called the *interchange fee*, and the specific percentage is determined not by the issuing bank, but by the card network, in this case Visa. The interchange fee is important because it accounts for the bulk of the merchant discount, and because it is a major source of revenue for card issuers. (The other major source of revenue is the interest payments from cardholders who keep a credit balance.)

Visa will then make a payment to the acquiring bank, but it will also keep a certain percentage of the \$100 for itself (H). That percentage, usually around 0.2%, represents the fees Visa collects for processing the transaction. (The processing fee includes both a fixed and variable component, so 0.2% is just a rough average figure.) Lastly, the acquiring bank pays the merchant, but also withholds a certain percentage of the \$100 (usually around 0.2%) as payment for its work in servicing the merchant (I). At the end of the month you will receive a billing statement that will include the \$100 purchase (J), and to keep your account in good standing, you will make a payment to the issuer (K).

So what does the merchant get in the end? Based on the percentages shown in Figure 9, it receives \$100.00 − \$1.70 − \$0.20 − \$0.20 = \$97.90, i.e., the transaction costs the merchant 2.1% of the original \$100. That 2.1% is called the merchant discount. The calculation is summarized in Figure 10.

Suppose that you buy your \$100 item using an American Express card. Because Amex is a closed system, the process is much simpler. Once again an authorization request is made, now to Amex (the card issuer). Assuming your account is in good standing, the authorization is approved. Now the merchant's request for payment goes directly to Amex. Amex will then pay the merchant, but retain a fee, which might be about 3%. That fee is the entire merchant discount. This is summarized in Figure 11.

6.4 Recent Litigation and Regulation

For many years, merchants have complained about the discounts that they pay when accepting a credit card, charge card, or debit card. In addition, some of the rules that the card networks have imposed on merchants and on issuing banks have been criticized as unfair

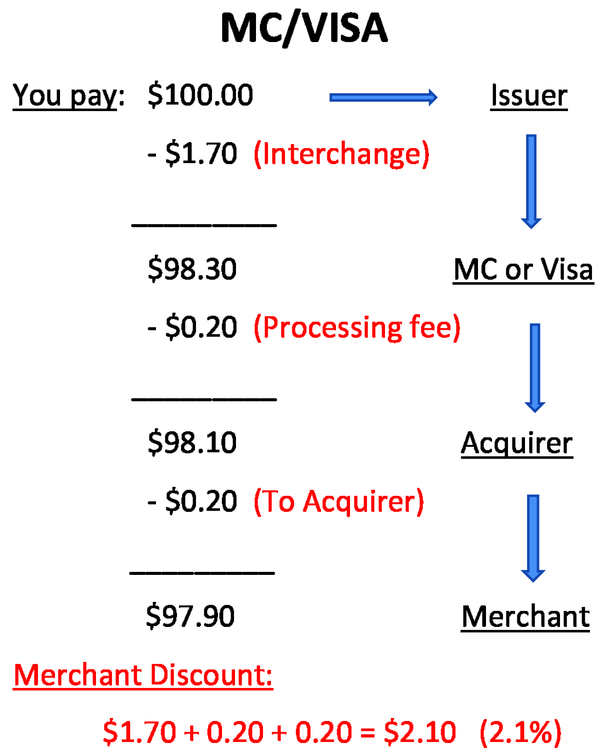

Figure 10: Merchant Discount: MC and Visa

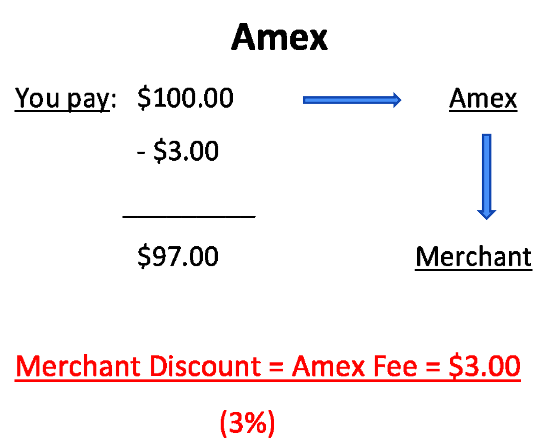

Figure 11: Merchant Discount: American Express

and/or anticompetitive. As a result, there has been a great deal of legal and regulatory activity over the past decade or so regarding card networks. Here are some of the more important results:

- **2000:** Previously banks that issued MC cards could also issue Visa cards, and vice versa. However they were prohibited from issuing Amex or Discover cards. The DOJ sued, and MC and Visa were forced to end their issuance restrictions.[7]

- **2005:** A large group of merchants sued MC and Visa over the discount on debit card transactions, and the requirement of MC and Visa that if they accepted the network's credit card, they also had to accept the debit card, and vice versa. As a settlement in that suit, MC and Visa were forced to end their "Honor all Cards" rule. They also agreed to lower their off-line debit interchange rates.

- **2007:** Amex sued MC and Visa for damages resulting from past issuance restrictions. MC and Visa agreed to settle (for about $1 billion and $2 billion respectively).

- **2010:** The DOJ sued MC, Visa, and Amex on their "steering" rules, which prohibit a merchant from offering a discount to a customer who pays in cash rather than with a card. MC and Visa agreed to stop; Amex decided to fight this in court.

- **2011:** The Federal Reserve reduced debit interchange rates further, and also considered regulating credit interchange. The Fed also issued restrictions on issuer practices.

- **2012:** MasterCard, Visa, and several major banks (including JPMorgan Chase and Bank of America) had been sued by a large group of merchants who claimed they acted anticompetitively in setting interchange rates and processing fees. In a settlement announced in July 2012, MC, Visa, and the banks agreed to pay the merchants more than $6 billion in damages, and also agreed to reduce their processing fees.

---

[7] For a detailed discussion of this litigation and the issues involved, see Robert S. Pindyck, "Governance, Issuance Restrictions, and Competition in Payment Card Networks," in J. Kwoka and L. White, Eds., *The Antitrust Revolution*, 6th Ed., Oxford Univ. Press.