Chapter Three

Types and Patterns of Innovation

Innovating in India: The chotuKool Project

Godrej & Boyce, founded in India in 1897, sold a range of products to the Indian market including household appliances, office furniture, and industrial process equipment. In recent years, international competitors such as Haier and Samsung were cutting deep into Godrej's market share for household appliances such as refrigerators, washing machines, and air conditioners, and management knew that to preserve the company would require innovative solutions.

One such solution was the chotuKool, a small, portable refrigerator. Though around the world refrigeration was considered a mature technology, in rural India as many as 90 percent of families could not afford household appliances, did not have reliable access to electricity, and had no means of refrigeration. This significantly limited the kinds of foods they could eat and how they could be prepared. Finding a way to provide refrigeration to this segment of the population offered the promise of both a huge market and making a meaningful difference in people's quality of life. As noted by Navroze Godrej, Directed of Special Projects at Godrej, "We imagined we would be making a shrunken down version of a refrigerator. Make it smaller, make it cheaper. And we had preconceived notions of how to build a brand that resonated with these users through big promotions and fancy ad campaigns."

These assumptions would turn out to be wrong. First, as Godrej's team looked at the options of how to reduce the cost of a conventional compressor-based refrigerator, they quickly realized that they could not reduce its cost by enough to make a meaningful difference.[a] Second, they discovered that having the refrigerator be lightweight was more important than they had previously thought because many rural Indians lived migratory lives, moving to follow the availability of work. Third, because of the lack of refrigeration, most people were in the habit of cooking just enough for the day, and thus had relatively low refrigeration capacity needs. Fourth, of those few rural Indians that did have refrigerators, many did not plug them in for most of the day for fear of them being damaged by power surges. As Godrej notes, "We were surprised by many things, we were shocked by many things . . . we realized our original hypothesis was quite wrong."[b]

Based on these insights, the company designed a small and portable refrigerator based on thermoelectric cooling (rather than compressor technology). Thermoelectric cooling was the cooling method used in laptops; it involved running a current between two semiconductors. It was far more expensive on a per-unit-of-cooling basis, but it had much lower power requirements and could be used on a much smaller scale than compressor cooling. This enabled Godrej to make a very small, lightweight refrigerator with a relatively low price (35–40 percent cheaper than traditional refrigerators). It also lowered the power costs of operating a refrigerator, and made the refrigerator able to operate for several hours on a 12-volt battery, making it much more adaptable to situations where power was unreliable.

In Godrej's initial plan for the chotuKool, the refrigerators would be cherry red and look like coolers. Soon, however, managers at chotuKool realized that if the refrigerators were just perceived as inexpensive alternatives to refrigerators, they had the potential to be stigmatizing for consumers who, in turn, would not talk about them to their friends. This was a serious problem because the company had counted on word of mouth to spread information about the refrigerators deep into rural communities. To get people to talk about the coolers they needed to be aspirational—they needed to be *cool*.

Godrej decided to revamp the design of the coolers, giving them a more sophisticated shape and making them customizable (buyers could choose from over 100 decorative skin colors for the chotuKool).[c] They also decided to market the refrigerators to the urban affluent market in addition to the rural market, as adoption by the urban affluent market would remove any stigma associated with buying them. To attract this market they positioned the refrigerators as perfect for picnics, parties, offices, dorm rooms, use in cars, and so on.

To get the chotuKool to rural customers would require a dramatically different distribution system than Godrej had traditionally used. However, building out a distribution system into rural communities would prohibitively raise the cost of chotuKool, potentially rendering the product nonviable. The development team was initially stumped. Then one day G. Sunderraman, vice president of Godrej and leader of the chotuKool project, happened to inquire with a university official about obtaining college application forms for his youngest son and the official pointed out that Sunderraman could get the forms at any post office. At that moment, Sunderraman realized that the post office, which had offices in every rural area of India, could be an ideal distribution channel for the chotuKool.[d] It was a very novel proposition, but India Post agreed to the collaboration and soon chotuKools were available in all post offices in the central region of India.[e] As Sunderraman noted, "The India Post network is very well spread in India and is about three or four times larger than the best logistic suppliers."[f]

The chotuKool won several design awards in its first years, and after selling 100,000 units in its second year *Fast Company* gave Godrej its "Most Innovative Company" award. Godrej and Sunderraman were disappointed to discover that it was not as rapidly adopted by rural poor households as they had hoped; the roughly $50 price was still too expensive for most poor rural families in India. However, the chotuKool turned out to be much more popular than anticipated among hotels, food stalls, flower shops, and other small stores because it enabled these small stores to offer higher valued products (such as cold drinks) or to keep products fresh longer, thereby increasing their profits. The chotuKool also became a popular lifestyle product among the urban affluent population who began to widely use them in their cars.

Godrej's experience developing and launching the chotuKool had provided many lessons. They had learned that to radically reduce the cost of a product might require completely rethinking the technology—sometimes even in ways that initially seemed more expensive. They learned that customers who had adapted their way of life to the lack of a technology (like refrigeration) might not adopt that technology even if it was made markedly less expensive. Finally, they learned not to underestimate the value of making a product work for multiple market segments, including those that might not be initially obvious as customers. Though some people considered chotuKool a failure because it had not achieved its original objective of wide adoption by the rural poor, Godrej (and many others) considered it a success: the product expanded Godrej's market share, penetrated new market segments in which Godrej had not formerly competed, and demonstrated Godrej's innovative capabilities to the world.

Discussion Questions

1. What were the pros and cons of attempting to develop a refrigerator for India's rural poor?
2. What product and process innovations did the chotuKool entail? Would you consider these incremental or radical? Architectural or component? Competence enhancing or competence destroying?
3. Did the chotuKool pose a threat of disrupting the traditional refrigerator market? Why or why not?
4. Is there anything you think Godrej should have done differently to penetrate the market of rural poor families in India?
5. What other products might the lessons Godrej learned with chotuKool apply to?

[a] McDonald, R., D. van Bever, and E. Ojomo, "chotuKool: 'Little Cool,' Big Opportunity," *Harvard Business School Case 616–020* (June 2016), revised September 2016.

[b] Furr, N., and J. Dyer, "How Godrej Became an Innovation Star," *Forbes* (May 13, 2015).

[c] www.chotukool.com, accessed June 26, 2018.

[d] Furr, N., and J. Dyer, "How Godrej Became an Innovation Star," *Forbes* (May 13, 2015).

[e] Nadu, T., "chotuKool Offer in Post Offices," *The Hindu* (June 9, 2013).

[f] "chotuKool: Keeping Things Cool with Frugal Innovation," *WIPO Magazine,* (December 2013).

OVERVIEW

The previous chapters pointed out that technological innovation can come from many sources and take many forms. Different types of technological innovations offer different opportunities for organizations and society, and they pose different demands upon producers, users, and regulators. While there is no single agreed-upon taxonomy to describe different kinds of technological innovations, in this chapter we will review several dimensions that are often used to categorize technologies. These dimensions are useful for understanding some key ways that one innovation may differ from another.

technology trajectory
The path a technology takes through its lifetime. This path may refer to its rate of performance improvement, its rate of diffusion, or other change of interest.

The path a technology follows through time is termed its **technology trajectory**. Technology trajectories are most often used to represent the technology's rate of performance improvement or its rate of adoption in the marketplace. Though many factors can influence these technology trajectories (as discussed in both this chapter and the following chapters), some patterns have been consistently identified in technology trajectories across many industry contexts and over many periods. Understanding these patterns of technological innovation provides a useful foundation that we will build upon in the later chapters on formulating technology strategy.

The chapter begins by reviewing the dimensions used to distinguish types of innovations. It then describes the s-curve patterns so often observed in both the rate of technology improvement and the rate of technology diffusion to the market. In the last section, the chapter describes research suggesting that technological innovation follows a cyclical pattern composed of distinct and reliably occurring phases.

TYPES OF INNOVATION

Technological innovations are often described using dimensions such as *radical* versus *incremental*. Different types of innovation require different kinds of underlying knowledge and have different impacts on the industry's competitors and customers. Four of the dimensions most commonly used to categorize innovations are described here: product versus process innovation, radical versus incremental, competence enhancing versus competence destroying, and architectural versus component.

Product Innovation versus Process Innovation

Product innovations are embodied in the outputs of an organization—its goods or services, even if those products are services. For example, Snapchat's filters and special effects that enable users to augment their photos are product innovations. Process innovations are innovations in the way an organization conducts its business, such as in the techniques of producing or marketing goods or services. For example, Elon Musk's use of automation for most of the production process for the Model 3 with giant robots is a process innovation. Process innovations are often oriented toward improving the effectiveness or efficiency of production by, for example, reducing defect rates or increasing the quantity that may be produced in a given time. For example, a process innovation at a biotechnology firm might entail developing a genetic algorithm that can quickly search a set of disease-related genes to identify a target for therapeutic intervention. In this instance, the process innovation (the genetic algorithm) can speed up the firm's ability to develop a product innovation (a new therapeutic drug).

New product innovations and process innovations often occur in tandem. First, new processes may enable the production of new products. For example, as discussed later in the chapter, the development of new metallurgical techniques enabled the development of the bicycle chain, which in turn enabled the development of multiple-gear bicycles. Second, new products may enable the development of new processes. For example, the development of advanced workstations has enabled firms to implement computer-aided manufacturing processes that increase the speed and efficiency of production. Finally, a product innovation for one firm may simultaneously be a process innovation for another. For example, when United Parcel Service (UPS) helps a customer develop a more efficient distribution system, the new distribution system is simultaneously a product innovation for UPS and a process innovation for its customer.

Though product innovations are often more visible than process innovations, both are extremely important to an organization's ability to compete. Throughout the remainder of the book, the term *innovation* will be used to refer to both product and process innovations.

Radical Innovation versus Incremental Innovation

radical innovation
An innovation that is very new and different from prior solutions.

incremental innovation
An innovation that makes a relatively minor change from (or adjustment to) existing practices.

One of the primary dimensions used to distinguish types of innovation is the continuum between radical versus incremental innovation. A number of definitions have been posed for **radical innovation** and **incremental innovation**, but most hinge on the degree to which an innovation represents a departure from existing practices.[1] Thus, radicalness might be conceived as the combination of *newness* and the degree of *differentness.* A technology could be new to the world, new to an industry, new to a firm, or new merely to an adopting business unit. A technology could be significantly different from existing products and processes or only marginally different. The most radical innovations would be new to the world and exceptionally different from existing products and processes. The introduction of wireless telecommunication products aptly illustrates this—it embodied significantly new technologies that required new manufacturing and service processes. Incremental innovation is at the other end of the spectrum. An incremental innovation might not be particularly new or exceptional; it might have been previously known to the firm or industry, and involve only a minor change from (or adjustment to) existing practices. For example, changing the screen of a cell phone to make it more crack resistant or offering a new service plan with better international texting rates would represent incremental innovation.

The radicalness of innovation is also sometimes defined in terms of risk. Since radical innovations often embody new knowledge, producers and customers will vary in their experience and familiarity with the innovation, and in their judgment of its usefulness or reliability.[2] The development of third generation (3G) telephony is illustrative. 3G wireless communication technology utilizes broadband channels. This increased bandwidth gave mobile phones far greater data transmission capabilities that enabled activities such as videoconferencing and accessing the most advanced Internet sites. For companies to develop and offer 3G wireless telecommunications service required a significant investment in new networking equipment and an infrastructure capable of carrying a much larger bandwidth of signals. It also required developing phones with greater display and memory capabilities, and either increasing the phone's battery power or increasing the efficiency of the phone's power utilization. Any of these technologies could potentially pose serious obstacles. It was also unknown to what degree customers would ultimately value broadband capability in a wireless device. Thus, the move to 3G required managers to assess several different risks simultaneously, including technical feasibility, reliability, costs, and demand.

Finally, the radicalness of an innovation is relative, and may change over time or with respect to different observers. An innovation that was once considered radical may eventually be considered incremental as the knowledge base underlying the innovation becomes more common. For example, while the first steam engine was a monumental innovation, today its construction seems relatively simple. Furthermore, an innovation that is radical to one firm may seem incremental to another. Although both Kodak and Sony introduced digital cameras for the consumer market within a year of each other (Kodak's DC40 was introduced in 1995, and Sony's Cyber-Shot Digital Still Camera was introduced in 1996), the two companies' paths to the introduction were quite different. Kodak's historical competencies and reputation were based on its expertise in chemical photography, and thus the transition to digital photography and video required a significant redirection for the firm. Sony, on the other hand, had been an electronics company since its inception, and had a substantial level of expertise in digital recording and graphics before producing a digital camera. Thus, for Sony, a digital camera was a straightforward extension of its existing competencies.

competence-enhancing and competence destroying innovation
A competence-enhancing innovation builds on existing knowledge and skills whereas a competence-destroying innovation renders existing knowledge and skills obsolete. Whether an innovation is competence enhancing or competence destroying depends on whose perspective is being taken. An innovation can be competence enhancing to one firm, while competence destroying for another.

Competence-Enhancing Innovation versus Competence-Destroying Innovation

Innovations can also be classified as **competence enhancing** versus **competence destroying**. An innovation is considered to be competence enhancing from the perspective of a particular firm if it builds on the firm's existing knowledge base. For example, each generation of Intel's microprocessors (e.g., 286, 386, 486, Pentium, Pentium II, Pentium III, Pentium 4) builds on the technology underlying the previous generation. Thus, while each generation embodies innovation, these innovations leverage Intel's existing competencies, making them more valuable.

An innovation is considered to be competence destroying from the perspective of a particular firm if the technology does not build on the firm's existing competencies or renders them obsolete. For example, from the 1600s to the early 1970s, no self-respecting mathematician or engineer would have been caught without a slide rule. Slide rules are lightweight devices, often constructed of wood, that use logarithm scales to solve complex mathematical functions. They were used to calculate everything from the structural properties of a bridge to the range and fuel use of an aircraft. Specially designed slide rules for businesses had, for example, scales for doing loan calculations or determining optimal purchase quantities. During the 1950s and 1960s, Keuffel & Esser was the preeminent slide-rule maker in the United States, producing 5000 slide rules a month. However, in the early 1970s, a new innovation relegated the slide rule to collectors and museum displays within just a few years: the inexpensive handheld calculator. Keuffel & Esser had no background in the electronic components that made electronic calculators possible and was unable to transition to the new technology. By 1976, Keuffel & Esser withdrew from the market.[3] Whereas the inexpensive handheld calculator built on the existing competencies of companies such as Hewlett-Packard and Texas Instruments (and thus for them would be competence enhancing), for Keuffel & Esser, the calculator was a competence-destroying innovation.

Architectural Innovation versus Component Innovation

Most products and processes are hierarchically nested systems, meaning that at any unit of analysis, the entity is a system of components, and each of those components is, in turn, a system of finer components, until we reach some point at which the components are elementary particles.[4] For example, a bicycle is a system of components such as a frame, wheels, tires, seat, brakes, and so on. Each of those components is also a system of components: The seat might be a system of components that includes a metal and plastic frame, padding, a nylon cover, and so on.

An innovation may entail a change to individual components, to the overall architecture within which those components operate, or both. An innovation is considered a **component innovation** (or **modular innovation**) if it entails changes to one or more components, but does not significantly affect the overall configuration of the system.[5] In the example above, an innovation in bicycle seat technology (such as the incorporation of gel-filled material for additional cushioning) does not require any changes in the rest of the bicycle architecture.

component (or modular) innovation
An innovation to one or more components that does not significantly affect the overall configuration of the system.

In contrast, an **architectural innovation** entails changing the overall design of the system or the way that components interact with each other. An innovation that is strictly architectural may reconfigure the way that components link together in the system, without changing the components themselves.[6] Most architectural innovations, however, create changes in the system that reverberate throughout its design, requiring changes in the underlying components in addition to changes in the ways those components interact. Architectural innovations often have far-reaching and complex influences on industry competitors and technology users.

architectural innovation
An innovation that changes the overall design of a system or the way its components interact with each other.

For example, the transition from the high-wheel bicycle to the safety bicycle was an architectural innovation that required (and enabled) the change of many components of the bicycle and the way in which riders propelled themselves. In the 1800s, bicycles had extremely large front wheels. Because there were no gears, the size of the front wheel directly determined the speed of the bicycle since the circumference of the wheel was the distance that could be traveled in a single rotation of the pedals. However, by the start of the twentieth century, improvements in metallurgy had enabled the production of a fine chain and a sprocket that was small enough and light enough for a human to power. This enabled bicycles to be built with two equally sized wheels, while using gears to accomplish the speeds that the large front wheel had enabled. Because smaller wheels meant shorter shock-absorbing spokes, the move to smaller wheels also prompted the development of suspension systems and pneumatic (air-filled) tires. The new bicycles were lighter, cheaper, and more flexible. This architectural innovation led to the rise of companies such as Dunlop (which invented the pneumatic tire) and Raleigh (which pioneered the three-speed, all-steel bicycle), and transformed the bicycle from a curiosity into a practical transportation device.

For a firm to initiate or adopt a component innovation may require that the firm have knowledge only about that component. However, for a firm to initiate or adopt an architectural innovation typically requires that the firm have architectural knowledge about the way components link and integrate to form the whole system. Firms must be able to understand how the attributes of components interact, and how changes in some system features might trigger the need for changes in many other design features of the overall system or the individual components. Modularity, and its role in the creation of platform ecosystems, is discussed in greater detail in Chapter Four.

Using the Dimensions

Though the dimensions described above are useful for exploring key ways that one innovation may differ from another, these dimensions are not independent, nor do they offer a straightforward system for categorizing innovations in a precise and consistent manner. Each of the above dimensions shares relationships with others—for example, architectural innovations are often considered more radical and more competence destroying than component innovations. Furthermore, how an innovation is described on a dimension often depends on who is doing the describing and with what it is being compared. An all-electric vehicle, for example, might seem like a radical and competence destroying innovation to a manufacturer of internal combustion engines, but to a customer who only has to change how they fuel/charge the vehicle, it might seem like an incremental and competence-enhancing innovation. Thus, while the dimensions above are valuable for understanding innovation, they should be considered relative dimensions whose meaning is dependent on the context in which they are used.

We now will turn to exploring patterns in technological innovation. Numerous studies of innovation have revealed recurring patterns in how new technologies emerge, evolve, are adopted, and are displaced by other technologies. We begin by examining technology s-curves.

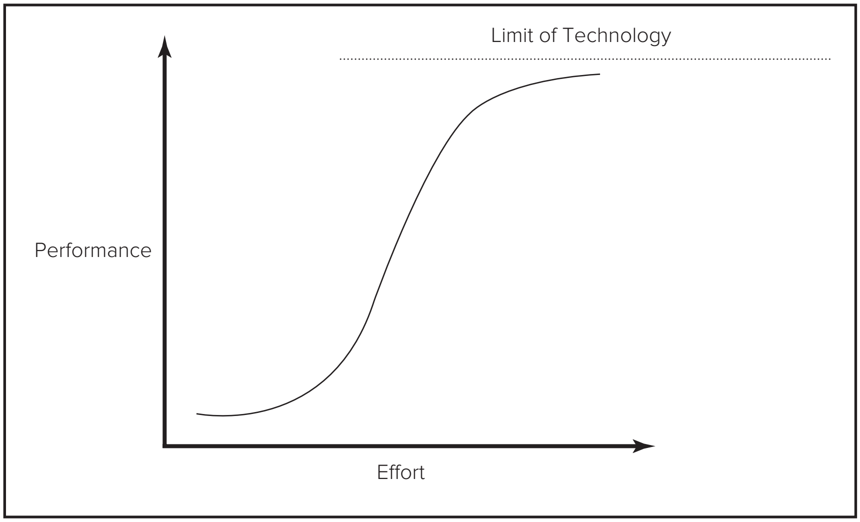
TECHNOLOGY S-CURVES

Both the rate of a technology's performance improvement and the rate at which the technology is adopted in the marketplace repeatedly have been shown to conform to an s-shape curve. Though s-curves in technology performance and s-curves in technology diffusion are related (improvements in performance may foster faster adoption, and greater adoption may motivate further investment in improving performance), they are fundamentally different processes. S-curves in technology improvement are described first, followed by s-curves in technology diffusion. This section also explains that despite the allure of using s-curves to predict when new phases of a technology's life cycle will begin, doing so can be misleading.

S-Curves in Technological Improvement

Many technologies exhibit an s-curve in their performance improvement over their lifetimes.[7] When a technology's performance is plotted against the amount of effort and money invested in the technology, it typically shows slow initial improvement, then

FIGURE 3.1
S-Curve of Technology Performance



accelerated improvement, then diminishing improvement (see Figure 3.1). Performance improvement in the early stages of a technology is slow because the fundamentals of the technology are poorly understood. Great effort may be spent exploring different paths of improvement or different drivers of the technology's improvement. If the technology is very different from previous technologies, there may be no evaluation routines that enable researchers to assess its progress or its potential. Furthermore, until the technology has established a degree of legitimacy, it may be difficult to attract other researchers to participate in its development.[8] However, as scientists or firms gain a deeper understanding of the technology, improvement begins to accelerate. The technology begins to gain legitimacy as a worthwhile endeavor, attracting other developers. Furthermore, measures for assessing the technology are developed, permitting researchers to target their attention toward those activities that reap the greatest improvement per unit of effort, enabling performance to increase rapidly. However, at some point, diminishing returns to effort begin to set in. As the technology begins to reach its inherent limits, the cost of each marginal improvement increases, and the s-curve flattens.

Often a technology's s-curve is plotted with performance (e.g., speed, capacity, or power) against time, but this must be approached with care. If the effort invested is not constant over time, the resulting s-curve can obscure the true relationship. If effort is relatively constant over time, plotting performance against time will result in the same characteristic curve as plotting performance against effort. However, if the amount of effort invested in a technology decreases or increases over time, the resulting curve could appear to flatten much more quickly, or not flatten at all. For instance, one of the more well-known technology trajectories is described by an axiom that became known as Moore's law. In 1965, Gordon Moore, cofounder of Intel, noted that the density of transistors on integrated circuits had doubled every year since the integrated circuit was invented. Figure 3.2 shows Intel's microprocessor transistor density from 1971 to 2007 and reveals a sharply increasing performance curve.

FIGURE 3.2
Improvements in Intel's Microprocessor Transistor Density over Time

Year	Transistors	Intel CPU
1971	2250	4004
1972	2500	8008
1974	5000	8080
1978	29,000	8086
1982	120,000	286
1985	275,000	386™
1989	1,180,000	486™ DX
1993	3,100,000	Pentium®
1997	7,500,000	Pentium II
1999	24,000,000	Pentium III
2000	42,000,000	Pentium 4
2002	55,000,000	Pentium M
2003	220,000,000	Itanium 2
2005	291,000,000	Pentium D
2006	582,000,000	Core 2 Quad
2007	731,000,000	Core i7 (Quad)

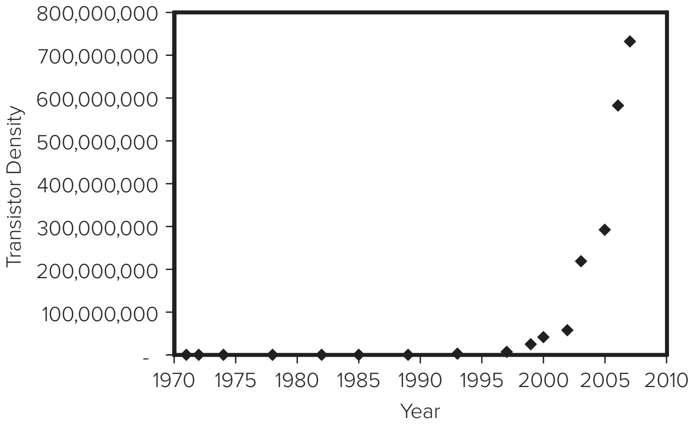
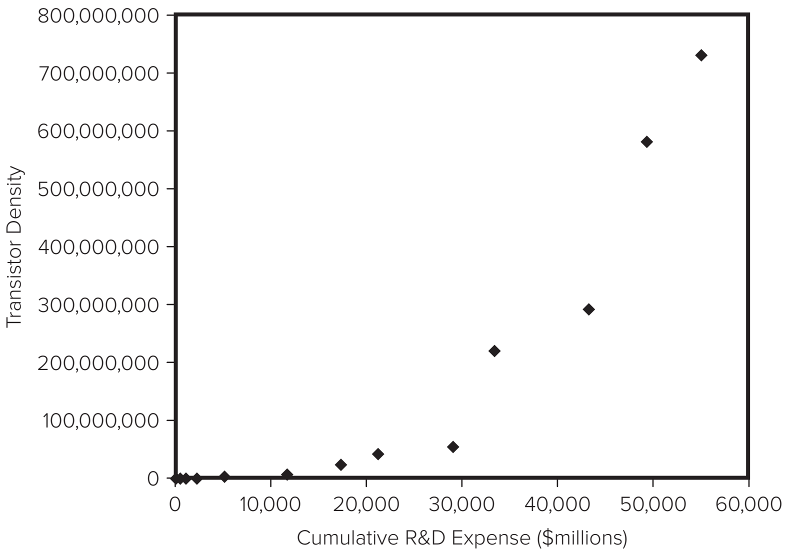


FIGURE 3.3
Graph of Transistor Density versus Cumulative R&D Expense, 1972–2007



However, Intel's rate of investment (research and development dollars per year) also increased rapidly over that time frame, as shown in Figure 3.3. Not all of Intel's R&D expense goes directly to improving microprocessor power, but it is reasonable to assume that Intel's investment specifically in microprocessors would exhibit a similar pattern of increase. Figure 3.3 shows that the big gains in transistor density have come at a big cost in terms of effort invested. Though the curve does not yet resemble the traditional s-curve, its rate of increase is not as sharp as when the curve is plotted against years.

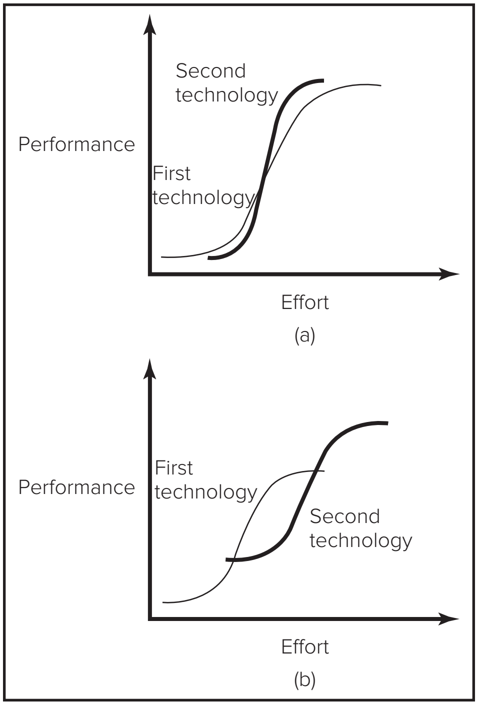
discontinuous technology
A technology that fulfills a similar market need by building on an entirely new knowledge base.

Technologies do not always get the opportunity to reach their limits; they may be rendered obsolete by new, **discontinuous technologies**. A new innovation is discontinuous when it fulfills a similar market need, but does so by building on an entirely new knowledge base.[9] For example, the switches from propeller-based planes to jets, from silver halide (chemical) photography to digital photography, from carbon copying to photocopying, and from audio on compact discs to MP3 were all technological discontinuities.

Initially, the technological discontinuity may have lower performance than the incumbent technology. For instance, one of the earliest automobiles, introduced in 1771 by Nicolas Joseph Cugnot, was never put into commercial production because it was much slower and harder to operate than a horse-drawn carriage. It was three-wheeled, steam-powered, and could travel at 2.3 miles per hour. A number of steam- and gas-powered vehicles were introduced in the 1800s, but it was not until the early 1900s that automobiles began to be produced in quantity.

In early stages, effort invested in a new technology may reap lower returns than effort invested in the current technology, and firms are often reluctant to switch. However, if the disruptive technology has a steeper s-curve (see Figure 3.4a) or an s-curve that increases to a higher performance limit (see Figure 3.4b), there may come a time when the returns to effort invested in the new technology are much higher than effort invested in the incumbent technology. New firms entering the industry are likely to choose the disruptive technology, and incumbent firms face the difficult choice of trying to extend the life of their current technology or investing in switching to the new technology. If the disruptive technology has much greater performance potential for a given amount of effort, in the long run it is likely to displace the incumbent technology, but the rate at which it does so can vary significantly.

FIGURE 3.4
Technology S-Curves—Introduction of Discontinuous Technology



S-Curves in Technology Diffusion

technology diffusion
The spread of a technology through a population.

S-curves are also often used to describe the diffusion of a technology. Unlike s-curves in technology performance, s-curves in **technology diffusion** are obtained by plotting the cumulative number of adopters of the technology against time. This yields an s-shape curve because adoption is initially slow when an unfamiliar technology is introduced to the market; it accelerates as the technology becomes better understood and utilized by the mass market, and eventually the market is saturated so the rate of new adoptions declines. For instance, when electronic calculators were introduced to the market, they were first adopted by the relatively small pool of scientists and engineers. This group had previously used slide rules. Then the calculator began to penetrate the larger markets of accountants and commercial users, followed by the still larger market that included students and the general public. After these markets had become saturated, fewer opportunities remained for new adoptions.[10]

One rather curious feature of technology diffusion is that it typically takes far more time than information diffusion.[11] For example, Mansfield found that it took 12 years for half the population of potential users to adopt industrial robots, even though these potential users were aware of the significant efficiency advantages the robots offered.[12] If a new technology is a significant improvement over existing solutions, why do some firms shift to it more slowly than others? The answer may lie in the complexity of the knowledge underlying new technologies, and in the development of complementary resources that make those technologies useful. Although some of the knowledge necessary to utilize a new technology might be transmitted through manuals or other documentation, other aspects of knowledge necessary to fully realize the potential of a technology might be built up only through experience. Some of the knowledge about the technology might be *tacit* and require transmission from person to person through extensive contact. Many potential adopters of a new technology will not adopt it until such knowledge is available to them, despite their awareness of the technology and its potential advantages.[13]

Furthermore, many technologies become valuable to a wide range of potential users only after a set of complementary resources are developed for them. For example, while the first electric light was invented in 1809 by Humphry Davy, an English chemist, it did not become practical until the development of bulbs within which the arc of light would be encased (first demonstrated by James Bowman Lindsay in 1835) and vacuum pumps to create a vacuum inside the bulb (the mercury vacuum pump was invented by Herman Sprengel in 1875). These early lightbulbs burned for only a few hours. Thomas Alva Edison built on the work of these earlier inventors when, in 1880, he invented filaments that would enable the light to burn for 1200 hours. The role of complementary resources and other factors influencing the diffusion of technological innovations are discussed further in Chapters four, five, and thirteen.

Finally, it should be clear that the s-curves of diffusion are in part a function of the s-curves in technology improvement: As technologies are better developed, they become more certain and useful to users, facilitating their adoption. Furthermore, as learning-curve and scale advantages accrue to the technology, the price of finished goods often drops, further accelerating adoption by users. For example, as shown in Figures 3.5 and 3.6, drops in average sales prices for video recorders, compact disc players, and cell phones roughly correspond to their increases in household penetration.

S-Curves as a Prescriptive Tool

Several authors have argued that managers can use the s-curve model as a tool for predicting when a technology will reach its limits and as a prescriptive guide for whether and when the firm should move to a new, more radical technology.[14] Firms can use data on the investment and performance of their own technologies, or data on the

FIGURE 3.5
Average Sales Prices of Consumer Electronics

Source: Consumer Electronics Association.

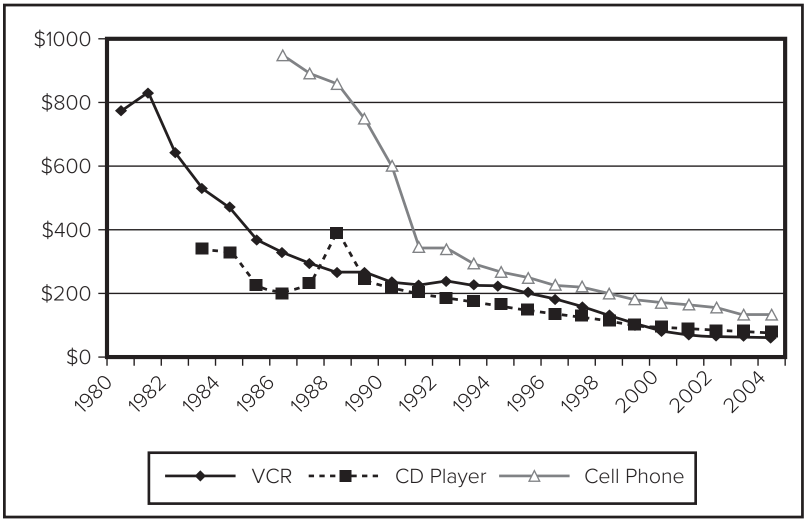
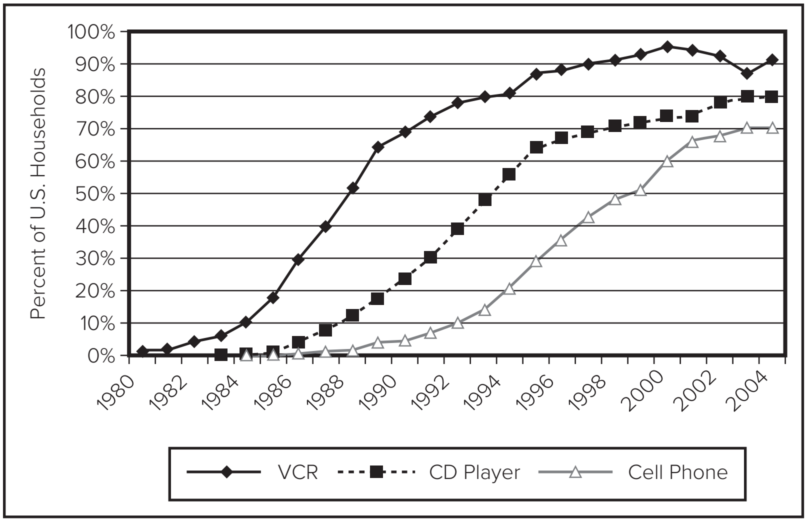


FIGURE 3.6
Penetration of Consumer Electronics

Source: Consumer Electronics Association.



overall industry investment in a technology and the average performance achieved by multiple producers. Managers could then use these curves to assess whether a technology appears to be approaching its limits or to identify new technologies that might be emerging on s-curves that will intersect the firm's technology s-curve. Managers could then switch s-curves by acquiring or developing the new technology. However, as a prescriptive tool, the s-curve model has several serious limitations.

Limitations of S-Curve Model as a Prescriptive Tool

First, it is rare that the true limits of a technology are known in advance, and there is often considerable disagreement among firms about what a technology's limits will be. Second, the shape of a technology's s-curve is not set in stone. Unexpected changes in the market, component technologies, or complementary technologies can shorten or extend the life cycle of a technology. Furthermore, firms can influence the shape of the s-curve through their development activities. For example, firms can sometimes stretch the s-curve through implementing new development approaches or revamping the architecture design of the technology.[15]

Christensen provides an example of this from the disk-drive industry. A disk drive's capacity is determined by its size multiplied by its areal recording density; thus, density has become the most pervasive measure of disk-drive performance. In 1979, IBM had reached what it perceived as a density limit of ferrite-oxide–based disk drives. It abandoned its ferrite-oxide–based disk drives and moved to developing thin-film technology, which had greater potential for increasing density. Hitachi and Fujitsu continued to ride the ferrite-oxide s-curve, ultimately achieving densities that were eight times greater than the density that IBM had perceived to be a limit.

Finally, whether switching to a new technology will benefit a firm depends on a number of factors, including (*a*) the advantages offered by the new technology, (*b*) the new technology's fit with the firm's current abilities (and thus the amount of effort that would be required to switch, and the time it would take to develop new competencies), (*c*) the new technology's fit with the firm's position in complementary resources (e.g., a firm may lack key complementary resources, or may earn a significant portion of its revenues from selling products compatible with the incumbent technology), and (*d*) the expected rate of diffusion of the new technology. Thus, a firm that follows an s-curve model too closely could end up switching technologies earlier or later than it should.

TECHNOLOGY CYCLES

The s-curve model above suggests that technological change is cyclical: Each new s-curve ushers in an initial period of turbulence, followed by rapid improvement, then diminishing returns, and ultimately is displaced by a new technological discontinuity.[16] The emergence of a new technological discontinuity can overturn the existing competitive structure of an industry, creating new leaders and new losers. Schumpeter called this process *creative destruction*, and argued that it was the key driver of progress in a capitalist society.[17]

Several studies have tried to identify and characterize the stages of the technology cycle in order to better understand why some technologies succeed and others fail, and whether established firms or new firms are more likely to be successful in introducing or adopting a new technology.[18] One technology evolution model that rose to prominence was proposed by Utterback and Abernathy. They observed that a technology passed through distinct phases. In the first phase (what they termed the *fluid phase*), there was considerable uncertainty about both the technology and its market. Products or services based on the technology might be crude, unreliable, or expensive, but might suit the needs of some market niches. In this phase, firms experiment with different form factors or product features to assess the market response. Eventually, however, producers and customers begin to arrive at some consensus about the desired

Research Brief The Diffusion of Innovation and Adopter Categories

S-curves in technology diffusion are often explained as a process of different categories of people adopting the technology at different times. One typology of adopter categories that gained prominence was proposed by Everett M. Rogers.[a] Figure 3.7 shows each of Rogers's adopter categories on a technology diffusion s-curve. The figure also shows that if the non cumulative share of each of these adopter groups is plotted on the vertical axis with time on the horizontal axis, the resulting curve is typically bell shaped (though in practice it may be skewed right or left).

INNOVATORS

Innovators are the first individuals to adopt an innovation. Extremely adventurous in their purchasing behavior, they are comfortable with a high degree of complexity and uncertainty. Innovators typically have access to substantial financial resources (and thus can afford the losses incurred in unsuccessful adoption decisions). Though they are not always well integrated into a particular social system, innovators play an extremely important role in the diffusion of an innovation because they are the individuals who bring new ideas into the social system. Rogers estimated that the first 2.5 percent of individuals to adopt a new technology are in this category.

EARLY ADOPTERS

The second category of adopters is the early adopters. Early adopters are well integrated into their social system and have the greatest potential for opinion leadership. Early adopters are respected by their peers and know that to retain that respect they must make sound innovation adoption decisions. Other potential adopters look to early adopters for information and advice; thus early adopters make excellent missionaries for new products or processes. Rogers estimated that the next 13.5 percent of individuals to adopt an innovation (after innovators) are in this category.

EARLY MAJORITY

Rogers identifies the next 34 percent of individuals in a social system to adopt a new innovation as the early majority. The early majority adopts innovations slightly before the average member of a social system. They are typically not opinion leaders, but they interact frequently with their peers.

LATE MAJORITY

The next 34 percent of the individuals in a social system to adopt an innovation are the late majority, according to Rogers. Like the early majority, the late majority constitutes one-third of the individuals in a social system. Those in the late majority approach innovation with a skeptical air and may not adopt the innovation until they feel pressure from their peers. The late majority may have scarce resources, thus making them reluctant to invest in adoption until most of the uncertainty about the innovation has been resolved.

LAGGARDS

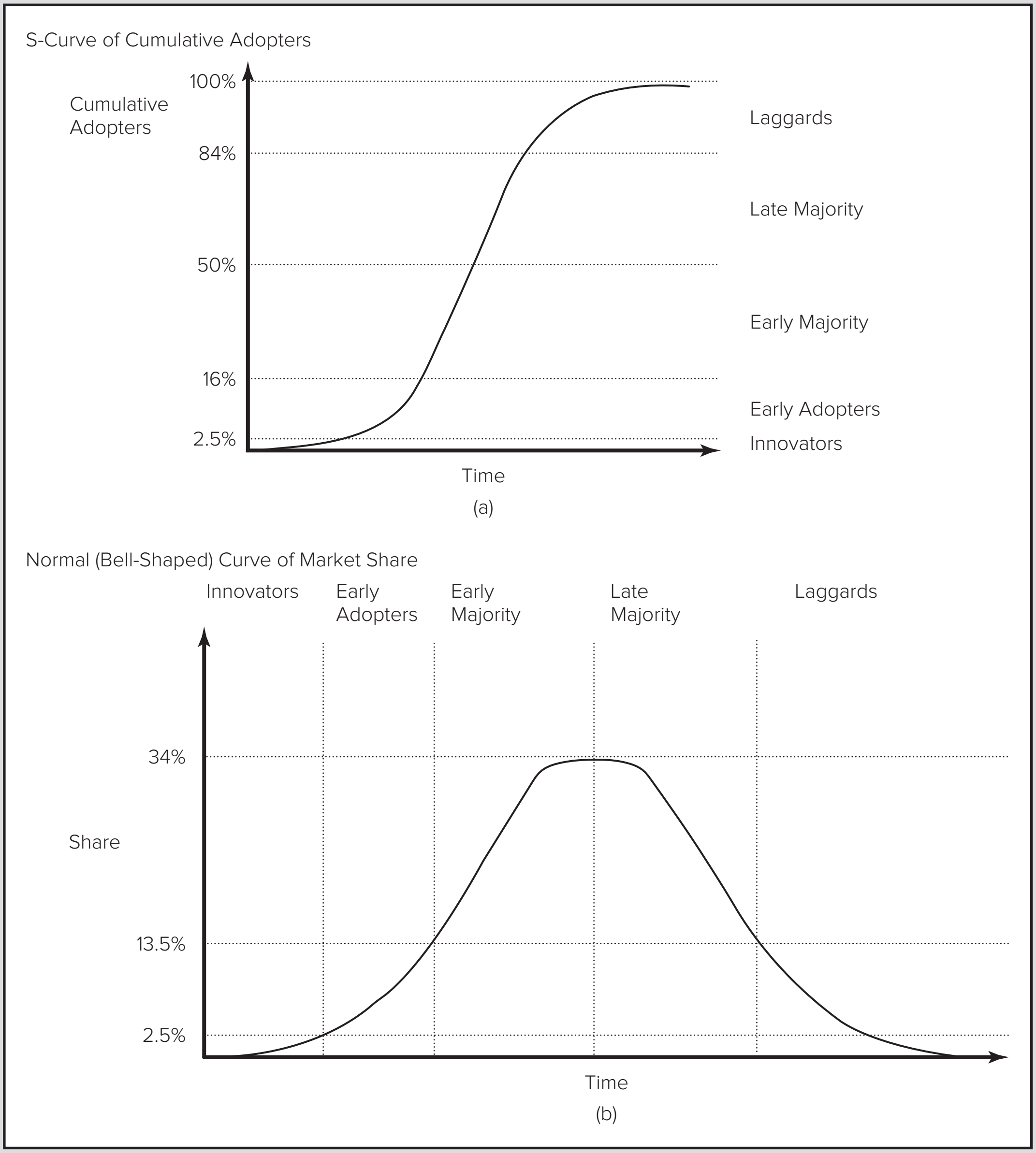
The last 16 percent of the individuals in a social system to adopt an innovation are termed *laggards*. They may base their decisions primarily upon past experience rather than influence from the social network, and they possess almost no opinion leadership. They are highly skeptical of innovations and innovators, and they must feel certain that a new innovation will not fail before adopting it.

[a] E. M. Rogers, *Diffusion of Innovations*, 5th ed. (New York: Free Press, 2003).

continued

concluded

FIGURE 3.7
Technology Diffusion S-Curve with Adopter Categories



From 1980 to 2012, Microsoft was entrenched as the dominant personal computer operating system, giving it enormous influence over many aspects of the computer hardware and software industries. Though competing operating systems had been introduced during that time (e.g., Unix, Geoworks, NeXTSTEP, Linux, and the Mac OS), Microsoft's share of the personal computer operating system market held stable at roughly 85 percent throughout most of that period. In 2013, however, Microsoft's dominance in computer operating systems was under greater threat than it had ever been. A high-stakes race for dominance over the next generation of computing was well underway, and Microsoft was not even in the front pack.

"SEGMENT ZERO"

As Andy Grove, former CEO of Intel, noted in 1998, in many industries—including microprocessors, software, motorcycles, and electric vehicles—technologies improve faster than customer demands of those technologies increase. Firms often add features (speed, power, etc.) to products faster than customers' capacity to absorb them. Why would firms provide higher performance than that required by the bulk of their customers? The answer appears to lie in the market segmentation and pricing objectives of a technology's providers. As competition in an industry drives prices and margins lower, firms often try to shift sales into progressively higher tiers of the market. In these tiers, high performance and feature-rich products can command higher margins. Though customers may also expect to have better-performing products over time, their ability to fully utilize such performance improvements is slowed by the need to learn how to use new features and adapt their work and lifestyles. Thus, while both the trajectory of technology improvement and the trajectory of customer demands are upward sloping, the trajectory for technology improvement is steeper (for simplicity, the technology trajectories are drawn in Figure 3.8 as straight lines and plotted against time in order to compare them against customer requirements).

In Figure 3.8, the technology trajectory begins at a point where it provides performance close to that demanded by the mass market, but over time it increases faster than the expectations of the mass market as the firm targets the high-end market. As the price of the technology rises, the mass market may feel it is overpaying for technological features it does not value. In Figure 3.9, the low-end market is not being served; it either pays far more for technology that it does not need or goes without. It is this market that Andy Grove, former CEO of Intel, refers to as segment zero.

For Intel, segment zero was the market for low-end personal computers (those less than $1000). While segment zero may seem unattractive in terms of margins, if it is neglected, it can become the breeding ground for companies that provide lower-end versions of the technology. As Grove notes, "The overlooked, underserved, and seemingly unprofitable end of the market can provide fertile ground for massive competitive change."[a]

FIGURE 3.8
Trajectories of Technology Improvement and Customer Requirements

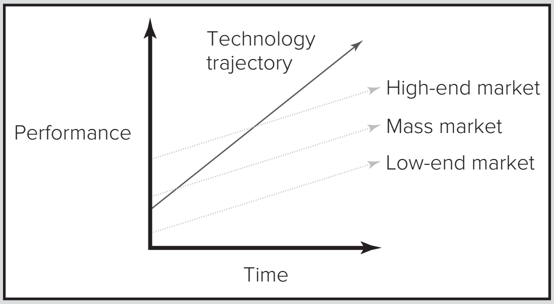
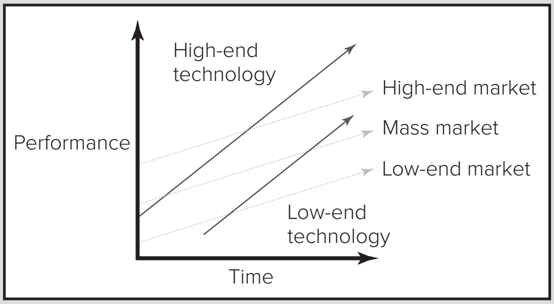


FIGURE 3.9
Low-End Technology's Trajectory Intersects Mass Market Trajectory



continued

concluded

As the firms serving low-end markets with simpler technologies ride up their own trajectories (which are also steeper than the slope of the trajectories of customer expectations), they can eventually reach a performance level that meets the demands of the mass market, while offering a much lower price than the premium technology (see Figure 3.9). At this point, the firms offering the premium technology may suddenly find they are losing the bulk of their sales revenue to industry contenders that do not look so low end anymore. For example, by 1998, the combination of rising microprocessor power and decreasing prices enabled personal computers priced under $1000 to capture 20 percent of the market.

THE THREAT TO MICROSOFT

So where was the "segment zero" that could threaten Microsoft? Look in your pocket. In 2018, Apple's iPhone operating system (iOS) and Google's Android collectively controlled almost 100 percent of the worldwide market for smartphones (with Android at 86.1 percent and iOS at 13.7 percent), followed by Research in Motion's Blackberry.[b] Gartner estimates put Microsoft's share at 3 percent. The iOS and Android interfaces offered a double whammy of beautiful aesthetics and remarkable ease of use. The applications business model used for the phones was also extremely attractive to both developers and customers, and quickly resulted in enormous libraries of applications that ranged from the ridiculous to the indispensible.

From a traditional economics perspective, the phone operating system market should not be that attractive to Microsoft—people do not spend as much on the applications, and the carriers have too much bargaining power, among other reasons. However, those smartphone operating systems soon became tablet operating systems, and tablets were rapidly becoming fully functional computers. Suddenly, all of that mindshare that Apple and Google had achieved in smartphone operating systems was transforming into mindshare in personal computer operating systems. Despite years of masterminding the computing industry, Microsoft's dominant position was at risk of evaporating.

[a] A. S. Grove, "Managing Segment Zero," *Leader to Leader,* 1999, p. 11.

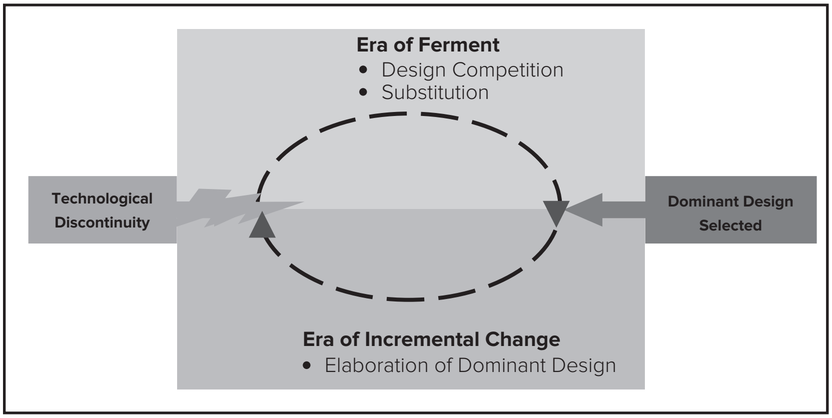
[b] www.Gartner.com, 2018.

dominant design
A product design that is adopted by the majority of producers, typically creating a stable architecture on which the industry can focus its efforts.

product attributes, and a **dominant design** emerges.[19] The dominant design establishes a stable architecture for the technology and enables firms to focus their efforts on process innovations that make production of the design more effective and efficient or on incremental innovations to improve components within the architecture. Utterback and Abernathy termed this phase the *specific phase* because innovations in products, materials, and manufacturing processes are all specific to the dominant design. For example, in the United States, the vast majority of energy production is based on the use of fossil fuels (e.g., oil, coal), and the methods of producing energy based on these fuels are well established. On the other hand, technologies that produce energy based on renewable resources (e.g., solar, wind, hydrogen) are still in the fluid phase. Organizations such as Royal Dutch/Shell, General Electric, and Ballard Power are experimenting with various forms of solar photocell technologies, wind-turbine technologies, and hydrogen fuel cells in efforts to find methods of using renewable resources that meet the capacity and cost requirements of serving large populations.

Building on the Utterback and Abernathy model, Anderson and Tushman studied the history of the U.S. minicomputer, cement, and glass industries through several cycles of technological change. Like Utterback and Abernathy, Anderson and Tushman found that each technological discontinuity inaugurated a period of turbulence and uncertainty (which they termed the *era of ferment*) (see Figure 3.10). The new technology might offer breakthrough capabilities, but there is little agreement about what the major subsystems of the technology should be or how they should be

FIGURE 3.10
The Technology Cycle



configured together. Furthermore, as later researchers noted, during the era of ferment different stakeholders might have different concepts of what purpose the technology should serve, or how a business model might be built around it.[20] Thus, while the new technology displaces the old (Anderson and Tushman refer to this as *substitution*), there is considerable design competition as firms experiment with different forms of the technology. Just as in the Utterback and Abernathy model, Anderson and Tushman found that a dominant design always arose to command the majority of the market share unless the next discontinuity arrived too soon and disrupted the cycle, or several producers patented their own proprietary technologies and refused to license to each other. Anderson and Tushman also found that the dominant design was never in the same form as the original discontinuity, but it was also never on the leading edge of the technology. Instead of maximizing performance on any individual dimension of the technology, the dominant design tended to bundle together a combination of features that best fulfilled the demands of the majority of the market.

In the words of Anderson and Tushman, the rise of a dominant design signals the transition from the era of ferment to the *era of incremental change*.[21] In this era, firms focus on efficiency and market penetration. Firms may attempt to achieve greater market segmentation by offering different models and price points. They may also attempt to lower production costs by simplifying the design or improving the production process. This period of accumulating small improvements may account for the bulk of the technological progress in an industry, and it continues until the next technological discontinuity.

Understanding the knowledge that firms develop during different eras lends insight into why successful firms often resist the transition to a new technology, even if it provides significant advantages. During the era of incremental change, many firms cease to invest in learning about alternative design architectures and instead invest in refining their competencies related to the dominant architecture. Most competition revolves around improving components rather than altering the architecture; thus, companies focus their efforts on developing component knowledge and knowledge related to the dominant architecture. As firms' routines and capabilities become more and more wedded to the dominant architecture, the firms become less able to identify and respond to a major architectural innovation. For example, the firm might establish divisions based on the primary components of the architecture and structure the communication channels between divisions on the basis of how those components interact. In the firm's effort to absorb and process the vast amount of information available to it, it is likely to establish filters that enable it to identify the information most crucial to its understanding of the existing technology design.[22] As the firm's expertise, structure, communication channels, and filters all become oriented around maximizing its ability to compete in the existing dominant design, they become barriers to the firm's recognizing and reacting to a new technology architecture.

While many industries appear to conform to this model in which a dominant design emerges, there are exceptions. In some industries, heterogeneity of products and production processes are a primary determinant of value, and thus a dominant design is undesirable.[23] For example, art and cuisine may be examples of industries in which there is more pressure to do things differently than to settle upon a standard.

Summary of Chapter

1. Different dimensions have been used to distinguish types of innovation. Some of the most widely used dimensions include product versus process innovation, radical versus incremental innovation, competence-enhancing versus competence-destroying innovation, and architectural versus component innovation.
2. A graph of technology performance over cumulative effort invested often exhibits an s-shape curve. This suggests that performance improvement in a new technology is initially difficult and costly, but, as the fundamental principles of the technology are worked out, it then begins to accelerate as the technology becomes better understood, and finally diminishing returns set in as the technology approaches its inherent limits.
3. A graph of a technology's market adoption over time also typically exhibits an s-shape curve. Initially the technology may seem uncertain and there may be great costs or risks for potential adopters. Gradually, the technology becomes more certain (and its costs may be driven down), enabling the technology to be adopted by larger market segments. Eventually the technology's diffusion slows as it reaches market saturation or is displaced by a newer technology.
4. The rate at which a technology improves over time is often faster than the rate at which customer requirements increase over time. This means technologies that initially met the demands of the mass market may eventually exceed the needs of the market. Furthermore, technologies that initially served only low-end customers (segment zero) may eventually meet the needs of the mass market and capture the market share that originally went to the higher-performing technology.
5. Technological change often follows a cyclical pattern. First, a technological discontinuity causes a period of turbulence and uncertainty, and producers and consumers explore the different possibilities enabled by the new technology. As producers and customers begin to converge on a consensus of the desired technological configuration, a dominant design emerges. The dominant design provides

a stable benchmark for the industry, enabling producers to turn their attention to increasing production efficiency and incremental product improvements. This cycle begins again with the next technological discontinuity.

6. The first design based on the initial technological discontinuity rarely becomes the dominant design. There is usually a period in which firms produce a variety of competing designs of the technology before one design emerges as dominant.
7. The dominant design rarely embodies the most advanced technological features available at the time of its emergence. It is instead the bundle of features that best meets the requirements of the majority of producers and customers.

Discussion Questions

1. What are some reasons that established firms might resist adopting a new technology?
2. Are well-established firms or new entrants more likely to (*a*) develop and/or (*b*) adopt new technologies? Why?
3. Think of an example of an innovation you have studied at work or school. How would you characterize it on the dimensions described at the beginning of the chapter?
4. What are some reasons that both technology improvement and technology diffusion exhibit s-shape curves?
5. Why do technologies often improve faster than customer requirements? What are the advantages and disadvantages to a firm of developing a technology beyond the current state of market needs?
6. In what industries would you expect to see particularly short technology cycles? In what industries would you expect to see particularly long technology cycles? What factors might influence the length of technology cycles in an industry?

Suggested Further Reading

Classics

Anderson, P., and M. L. Tushman, "Technological discontinuities and dominant designs," *Administrative Science Quarterly* 35 (1990), pp. 604–33.

Bijker, W. E., T. P. Hughes, and T. J. Pinch, *The Social Construction of Technological Systems* (Cambridge, MA: MIT Press, 1987).

Christensen, C. M., *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* (Boston: Harvard Business School Publishing, 1997).

Dosi, G., "Technological paradigms and technological trajectories," *Research Policy* 11 (1982), pp. 147–60.

Rogers, E., *Diffusion of Innovations,* 5th ed. (New York: Simon & Schuster Publishing, 2003).

Schilling, M. A., and M. Esmundo, "Technology S-Curves in Renewable Energy Alternatives: Analysis and Implications for Industry and Government," *Energy Policy,* 37 (2009), pp. 1767–81.

Recent Work

Ander, R., and R. Kapoor, "Innovation Ecosystems and the Pace of Substitution: Re-examining Technology S-curves," *Strategic Management Journal* (2015), 37:625–648.

Ethiraj, S., D. Levinthal, and R. R. Roy, "The dual role of modularity: Innovation and imitation," *Management Science,* 54 (2008), pp. 93–955.

Park, W., Y. K. Ro, and N. Kim, 2018. "Architectural Innovation and the Emergence of a Dominant Design: The Effects of Strategic Sourcing on Performance," 47 (2018), pp. 326–341.

Schilling, M. A., "Technology Shocks, Technological Collaboration, and Innovation Outcomes," *Organization Science,* 26 (2015), pp. 668–686.

Slater, S. F., J. J. Mohr, and S. Sengupta, "Radical Product Innovation Capability: Literature Review, Synthesis, and Illustrative Research Propositions," *Journal of Product Innovation Management,* 31 (2014), pp. 552–566.

Young, H. P., "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *American Economic Review* 99 (2009), pp. 1899–1924.

Endnotes

1. R. L. Daft and S. W. Becker, *Innovation in Organizations* (New York: Elsevier, 1978); T. D. Duchesneau, S. Cohn, and J. Dutton, *A Study of Innovation in Manufacturing: Determination, Processes and Methodological Issues,* vol. 1 (Social Science Research Institute, University of Maine, 1979); and J. Hage, *Theories of Organization* (New York: Wiley Interscience, 1980).
2. R. D. Dewar and J. E. Dutton, "The Adoption of Radical and Incremental Innovations: An Empirical Analysis," *Management Science* 32 (1986), pp. 1422–33; and J. Dutton and A. Thomas, "Relating Technological Change and Learning by Doing," in *Research on Technological Innovation, Management and Policy,* ed. R. Rosenbloom (Greenwich, CT: JAI Press, 1985), pp. 187–224.
3. C. Scuria-Fontana, "The Slide Rule Today: Respect for the Past; History of the Slide Rule," *Mechanical Engineering-CIME,* July 1990, pp. 122–24.
4. H. Simon, "The Architecture of Complexity," *Proceedings of the American Philosophical Society* 106 (1962), pp. 467–82.
5. L. Fleming and O. Sorenson, "Navigating the Technology Landscape of Innovation," *Sloan Management Review* 44, no. 2 (2003), p. 15; and M. A. Schilling, "Towards a General Modular Systems Theory and Its Application to Interfirm Product Modularity," *Academy of Management Review* 25 (2000), pp. 312–34.
6. R. Henderson and K. Clark, "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly* 35 (1990), pp. 9–30.
7. R. Foster, *Innovation: The Attacker's Advantage* (New York: Summit Books, 1986).
8. R. Garud and M. A. Rappa, "A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants," *Organization Science* 5 (1994), pp. 344–62; and W. E. Bijker, T. P. Hughes, and T. J. Pinch, *The Social Construction of Technological Systems* (Cambridge, MA: MIT Press, 1987).
9. Foster, *Innovation.*
10. R. Brown, "Managing the 's' Curves of Innovation," *Journal of Consumer Marketing* 9 (1992), pp. 61–72.

11. E. Rogers, *Diffusion of Innovations,* 4th ed. (New York: Free Press, 1995).
12. E. Mansfield, "Industrial Robots in Japan and the USA," *Research Policy* 18 (1989), pp. 183–92.
13. P. A. Geroski, "Models of Technology Diffusion," *Research Policy* 29 (2000), pp. 603–25.
14. Foster, *Innovation;* and E. H. Becker and L. M. Speltz, "Putting the S-curve Concept to Work," *Research Management* 26 (1983), pp. 31–33.
15. C. Christensen, *Innovation and the General Manager* (New York: Irwin/McGraw-Hill, 1999).
16. P. Anderson and M. Tushman, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change," *Administrative Science Quarterly* 35 (1990), pp. 604–34.
17. J. Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper Brothers, 1942).
18. See, for example, J. M. Utterback and W. J. Abernathy, "A Dynamic Model of Process and Product Innovation," *Omega, the International Journal of Management Science* 3 (1975), pp. 639–56; and D. Sahal, *Patterns of Technological Innovation* (Reading, MA: Addison-Wesley Publishing Co., 1981).
19. Utterback and Abernathy, "A Dynamic Model of Process and Product Innovation"; F. F. Suarez and J. M. Utterback, "Dominant Designs and the Survival of Firms," *Strategic Management Journal* 16 (1995), pp. 415–30; and J. M. Utterback and F. F. Suarez, "Innovation, Competition and Industry Structure," *Research Policy* 22 (1993), pp. 1–21.
20. Kaplan, S. and Tripsas, M. "Thinking about Technology: Applying a Cognitive Lens to Technical Change," *Research Policy,* 37 (2008):790–805.
21. P. Anderson and M. Tushman, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change," *Administrative Science Quarterly* 35 (1990), pp. 604–34.
22. R. Henderson and K. Clark, "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly* 35 (1990), pp. 9–30.
23. M. E. Porter, "The Technological Dimension of Competitive Strategy," in *Research on Technological Innovation, Management and Policy,* ed. R. S. Rosenbloom (Greenwich, CT: JAI Press, 1983); and S. Klepper, "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review* 86 (1996), pp. 562–83.

Chapter **Four**

Standards Battles, Modularity, and Platform Competition

A Battle for Dominance in Mobile Payments

By 2018, there were roughly five billion mobile broadband subscribers in the world.[a] As smartphones spread worldwide, so do mobile payment systems. The fastest growth is in developing economies in Asia, Africa, and Latin America, where many people did not have credit cards or bank cards and are transitioning directly from cash payments to mobile payments (see Figure 1).[b] It is difficult to get a precise picture of worldwide mobile payment system use and estimates vary widely, but they are all large: from hundreds of billions to trillions of U.S. dollars. However, in 2018, there was no dominant mobile payment system standard, and a battle among competing mobile payment mechanisms and standards was unfolding.

Many of the large mobile payment systems such as Apple Pay, Samsung Pay, or Android Pay, use Near Field Communication (NFC) chips in smartphones. NFC chips enable communication between a mobile device and a point-of-sale system just by having the devices in close proximity.[c] These systems transfer the customer's information wirelessly, and then merchant banks and credit card systems such as Visa or Mastercard complete the transaction. These systems are thus very much like existing ways of using credit cards, but enable completion of the purchase without contact. In emerging markets such as Asia-Pacific and Latin America, where NFC-enabled smartphones are less common, mobile payment systems are more likely to use QR codes (machine-readable bar codes), contactless stickers, and magnetic secure transmission (MST). MST sends a magnetic signal from a mobile device to a payment terminal.

The largest mobile payment system in the world, Alipay (owned by Alibaba group in China) uses a system based on QR codes. With Alipay, a merchant generates a barcode at the point of sale, and the consumer scans with their phone. An application then shows the details of the transaction and the

FIGURE 1
Proximity Mobile Payment Users Worldwide, by Region, 2016–2021

Source: eMarketer Report, 2018

	2016	2017	2018	2019	2020	2021
Proximity mobile payment users (millions)						
Asia-Pacific	413.4	569.9	650.7	722.6	793.8	855.6
North America	42.0	53.0	60.7	68.0	74.2	79.3
Western Europe	33.4	42.2	49.7	56.6	62.9	68.6
Central & Eastern Europe	17.3	23.3	38.8	33.2	37.4	41.7
Latin America	17.4	23.2	28.7	34.2	39.9	45.1
Middle East & Africa	7.0	9.6	12.8	16.5	20.7	24.8
Worldwide	**530.6**	**721.2**	**831.4**	**931.3**	**1028.9**	**1115.2**
Proximity mobile payment user growth (% change)						
Middle East & Africa	61.5%	37.5%	32.4%	29.8%	24.8%	20.0%
Latin America	60.2%	33.4%	23.8%	19.6%	16.1%	13.2%
Central & Eastern Europe	49.5%	34.6%	23.7%	15.3%	12.8%	11.5%
Western Europe	38.4%	26.3%	17.6%	13.9%	11.1%	9.1%
North America	49.3%	26.0%	14.6%	12.1%	9.1%	6.9%
Asia-Pacific	90.3%	37.8%	14.2%	11.1%	9.8%	7.8%
Worldwide	**79.1%**	**35.9%**	**15.3%**	**12.0%**	**10.5%**	**8.4%**

Note: ages 14+; mobile phone users who have made at least one proximity mobile payment transaction in the past 6 months; includes point-of-sale transactions made by using mobile devices as a payment method; excludes transactions made via tablet.

consumer enters a pin to confirm payment. Alipay reports that by the end of 2017 it had 520 million active users.

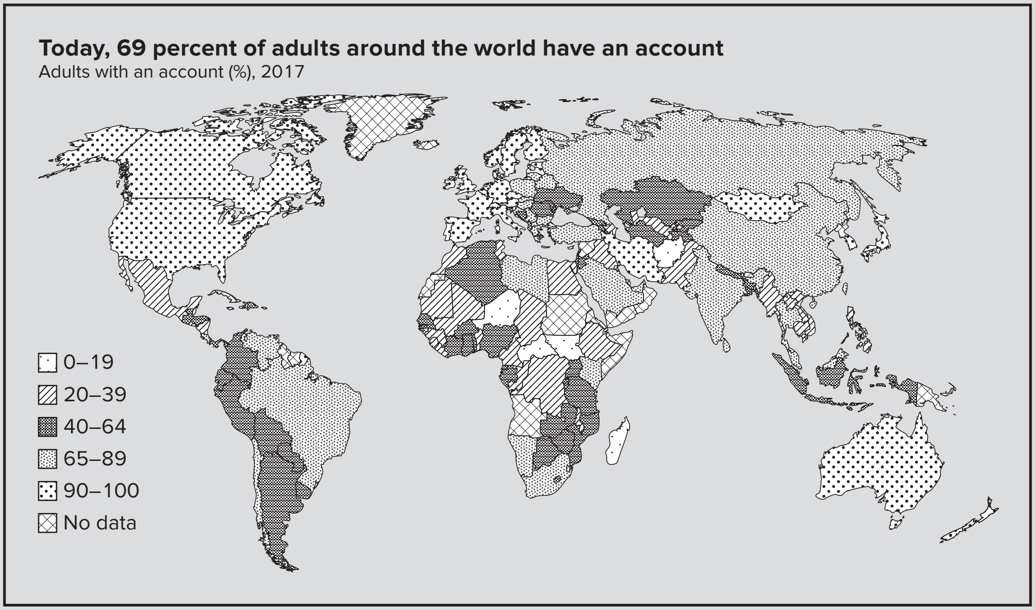
Other competitors, such as Square (with Square Wallet) and PayPal, use a downloadable application and the Web to transmit a customer's information. Square had gained early fame by offering small, free, credit card readers that could be plugged into the audio jack of a smartphone. These readers enabled vendors that would normally take only cash (street vendors, babysitters, etc.) to accept major credit cards.[d] Square processed $30 billion in payments in 2014, making the company one of the fastest growing tech start-ups in Silicon Valley.[e] Square takes about 2.75 to 3 percent from each transaction it processes, but must split that with credit card companies and other financial institutions. In terms of installed base, however, PayPal had the clear advantage, with more than 227 million active registered accounts by year end 2017. With PayPal, customers could complete purchases simply by entering their phone numbers and a pin number, or use a PayPal-issued magnetic stripe cards linked to their PayPal accounts. Users could opt to link their PayPal accounts to their credit cards, or directly to their bank accounts. PayPal also owned a service called Venmo that enabled peer-to-peer exchanges with a Facebook-like interface that was growing in popularity as a way to exchange money without carrying cash. Venmo charged a 3 percent fee if the transaction used a major credit card, but was free if the consumer used it with a major bank card or debit card.

In other parts of the world, intriguing alternatives for mobile banking are gaining traction quickly. In India and Africa, for example, there are enormous populations of "unbanked" or "underbanked" people (individuals who do not have bank accounts or make limited use of banking services). In these regions, the proportion of people with mobile phones vastly exceeds the proportion of people with credit cards. According to a GSMA report, for example, in sub-Saharan Africa, the number of mobile money accounts surpassed the number of bank accounts in 2015, and in 2016 more than 40 percent of the adult population of Kenya, Tanzania, Ghana, and Paraguay actively used mobile payment systems.[f]

The World Bank estimates that roughly two billion people worldwide do not have access to financial services, and 31 percent of adults have no bank account (Figure 2). This is a serious obstacle to overcoming poverty—access to banking is a very important resource for people to save money and utilize credit. Fortunately, the rise of mobile payment systems could have enormously beneficial social and economic consequences by helping the unbanked become banked.

FIGURE 2
Percent of Adults with a Bank Account

Source: Global Findex database.



In parts of Africa, where the proportion of people who are unbanked is very large, a system called M-Pesa ("M" for mobile and "pesa," which is kiswahili for money) enables any individual with a passport or national ID card to deposit money into his or her phone account, and transfer money to other users using Short Message Service (SMS).[g] By 2017, there were roughly 30 million M-Pesa users in 10 countries. The system had grown to offer a range of services including international transfers, loans, and health provision. It processed about six billion transactions in 2016, hitting a peak rate of 529 per second.[h]

As noted above, some of the mobile systems did not require involvement of the major credit card companies. PayPal, and its peer-to-peer system Venmo, for instance, did not require credit cards, nor does Alipay. A mobile payment system that cuts out the credit card companies could potentially save (or capture) billions of dollars in transaction fees. Credit card companies and merchants thus both had high incentives to influence the outcome of this battle.

For consumers, the key dimensions that influenced adoption were convenience (e.g., would the customer have to type in a code at the point of purchase? Was it easily accessible on a device the individual already owned?), risk of fraud (e.g., was the individual's identity and financial information at risk?), and ubiquity (e.g., could the system be used everywhere? Did it enable peer-to-peer transactions?). For merchants, the primary concerns were fraud and cost (e.g., what were the fixed costs and transaction fees of using the system?). Apple Pay had a significant convenience advantage in that a customer could pay with their fingerprint.[i] QR code–based systems, by contrast, required the customer to open the application on their phone and get a QR code that would need to be scanned at the checkout or to type in a pin.

By early 2018, it was clear that mobile payments represented a game-changing opportunity that could accelerate e-commerce, smartphone adoption, and the global reach of financial services. However, lack of compatibility between many of the mobile payment systems and uncertainty over what type of mobile payment system would become dominant still posed significant obstacles to consumer and merchant adoption, particularly in countries where most consumers already had credit cards.

Discussion Questions

1. What are some of the advantages and disadvantages of mobile payment systems in (a) developed countries and (b) developing countries?
2. What are the key factors that differentiate the different mobile payment systems? Which factors do consumers care most about? Which factors do merchants care most about?
3. Are there forces that are likely to encourage one of the mobile payment systems to emerge as dominant? If so, what do you think will determine which becomes dominant?
4. Is there anything the mobile payment systems could do to increase the likelihood of them becoming dominant?
5. How do these different mobile systems increase or decrease the power of (a) banks and (b) credit cards?

[a] International Telecommunications Union ICT Facts and Figures 2017.
[b] *World Payments Report 2017,* Capgemini and BNP Paribas, https://www.worldpaymentsreport.com.
[c] Kent, J., "Dominant Mobile Payment Approaches and Leading Mobile Payment Solution Providers: A Review," *Journal of Payments Strategy & Systems* 6, no. 4 (2012):315–24.
[d] Helft, M., "The Death of Cash," *Fortune* 166, no. 2 (2012):118–28.
[e] Isaac, M., "Square Expands Its Reach into Small-Business Services." *New York Times* (March 8, 2015).
[f] *State of the Industry Report on Mobile Money,* GSMA, https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/03/GSMA_State-of-the-Industry-Report-on-Mobile-Money_2016.pdf.
[g] Govindarajan, V., and M. Balakrishnan, "Developing Countries Are Revolutionizing Mobile Banking," *Harvard Business Review,"* Blog Network, (April 30, 2012).
[h] Monks, K., M-Pesa: Kenya's Mobile Money Success Story Turns 10. *CNN* (February 24, 2017).
[i] Pogue, D., "How Mobile Payments Are Failing and Credit Cards Are Getting Better," in Scientific American (January 20, 2015).

OVERVIEW

The previous chapter described recurrent patterns in technological innovation, and one of those patterns was the emergence of a dominant design. As Anderson and Tushman pointed out, the technology cycle almost invariably exhibits a stage in which the industry selects a **dominant design**. Once this design is selected, producers and customers focus their efforts on improving their efficiency in manufacturing, delivering, marketing, or deploying this dominant design, rather than continue to develop and consider alternative designs. In this chapter, we first will examine why industries experience strong pressure to select a single technology design as dominant and the multiple dimensions of value that will shape which technology designs rise to dominance. We will then look at why and how modularity and platform competition emerges in some industries.

WHY DOMINANT DESIGNS ARE SELECTED

dominant design
A single product or process architecture that dominates a product category—usually 50 percent or more of the market. A dominant design is a *de facto standard,* meaning that while it may not be officially enforced or acknowledged, it has become a standard for the industry.

Why do many markets coalesce around a single dominant design rather than support a variety of technological options? One primary reason is that many industries exhibit **increasing returns to adoption**, meaning that the more a technology is adopted, the more valuable it becomes.[1] Complex technologies often exhibit increasing returns to adoption in that the more they are used, the more they are improved. A technology that is adopted usually generates revenue that can be used to further develop and refine the technology. Furthermore, as the technology is used, greater knowledge and understanding of the technology accrue, which may then enable improvements both in the technology itself and in its applications. Finally, as a technology becomes more widely adopted, complementary assets are often developed that are specialized to operate with the technology. These effects can result in a self-reinforcing mechanism that increases the dominance of a technology regardless of its superiority or inferiority to competing technologies. Two of the primary sources of increasing returns are (1) learning effects and (2) network externalities.

increasing returns
When the *rate of return* (not just gross returns) from a product or process increases with the size of its installed base.

Learning Effects

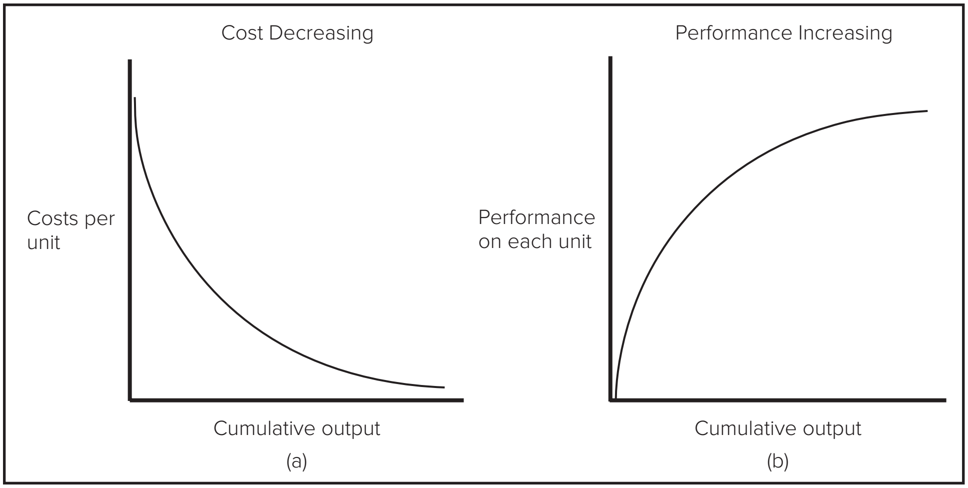
Ample empirical evidence shows that the more a technology is used, the more it is developed and the more effective and efficient it becomes.[2] As a technology is adopted, it generates sales revenues that can be reinvested in further developing and refining the technology. Furthermore, as firms accumulate experience with the technology, they find ways to use the technology more productively, including developing an organizational context that improves the implementation of the technology. Thus, the more a technology is adopted, the better it should become.

One example of learning effects is manifest in the impact of cumulative production on cost and productivity—otherwise known as the *learning curve*. As individuals and producers repeat a process, they learn to make it more efficient, often producing new technological solutions that may enable them to reduce input costs or waste rates. Organizational learning scholars typically model the learning curve as a function of cumulative output: performance increases, or cost decreases, with the number of units of production, usually at a decreasing rate (see Figure 4.1). For example, in studies of industries as diverse as aircraft production and pizza franchises, researchers have consistently found that the cost of producing a unit (e.g., a pizza or an airplane) falls as the number of units produced increases.

The standard form of the learning curve is formulated as $y = ax^{-b}$, where y is the number of direct labor hours required to produce the xth unit, a is the number of direct labor hours required to produce the first unit, x is the cumulative number of units produced, and b is the learning rate. This pattern has been found to be consistent with production data on a wide range of products and services, including the production of automobiles, ships, semiconductors, pharmaceuticals, and even heart surgery techniques.[3] Learning curves have also been identified by using a variety of performance measures, including productivity, total costs per unit, accidents per unit, and waste per unit.[4]

Though learning curves are found in a wide range of organizational processes, there are substantial differences in the rates at which organizations learn.[5] Both managers

FIGURE 4.1
Standard Learning-Curve Forms



and scholars are very interested in understanding why one firm reaps great improvement in a process while another exhibits almost no learning. Many studies have examined reasons for this variability, including looking at how the firm's learning rate is affected by process-improvement projects, intentional innovation, or contact with customers and suppliers.[6] The results suggest the learning rate can be influenced by factors such as the nature of the task, firm strategy, and the firm's prior experience.

Prior Learning and Absorptive Capacity

absorptive capacity
The ability of an organization to recognize, assimilate, and utilize new knowledge.

A firm's investment in prior learning can accelerate its rate of future learning by building the firm's absorptive capacity.[7] **Absorptive capacity** refers to the phenomenon whereby as firms accumulate knowledge, they also increase their future ability to assimilate information. A firm's prior related experience shapes its ability to recognize the value of new information, and to utilize that information effectively. For example, in developing a new technology, a firm will often try a number of unsuccessful configurations or techniques before finding a solution that works well. This experimentation builds a base of knowledge in the firm about how key components behave, what alternatives are more likely to be successful than others, what types of projects the firm is most successful at, and so on. This knowledge base enables the firm to more rapidly assess the value of related new materials, technologies, and methods. The effects of absorptive capacity suggest that firms that develop new technologies ahead of others may have an advantage in staying ahead. Firms that forgo investment in technology development may find it very difficult or expensive to develop technology in a subsequent period. This explains, in part, why firms that fall behind the technology frontier find it so difficult to catch up.

At the aggregate level, the more firms that are using a given technology and refining it, the more absorptive capacity that is being generated related to that technology, making development of that technology (and related technologies) more effective and efficient. Furthermore, as firms develop complementary technologies to improve the productivity or ease of utilization of the core technology, the technology becomes more attractive to other firms. In sum, learning effects suggest that early technology offerings often have an advantage because they have more time to develop and become enhanced than subsequent offerings. (However, as we shall discuss in Chapter Five, it is also possible to be *too early* to a market!)

Network Externalities

network externalities
Also termed *positive consumption externalities*, this is when the value of a good to a user increases with the number of other users of the same or similar good.

Many markets are characterized by **network externalities**, or positive consumption externalities.[8] In a market characterized by network externalities, the benefit from using a good increases with the number of other users of the same good. The classic examples of markets demonstrating network externality effects are those involving physical networks, such as railroads or telecommunications. Railroads are more valuable as the size of the railroad network (and therefore the number of available destinations) increases. Similarly, a telephone is not much use if only a few people can be called with it—the amount of utility the phone provides is directly related to the size of the network.

Network externalities can also arise in markets that do not have physical networks. For example, a user's benefit from using a good may increase with the number of users of the same good when compatibility is important. The number of users of a particular technology is often referred to as its **installed base**. A user may choose a computer platform based on the number of other users of that platform, rather than on the technological benefits of a particular platform, because it increases the ease of exchanging files. For example, many people choose a computer that uses the Windows operating system and an Intel microprocessor because the "Wintel" (*Windows* and *Intel*) platform has the largest installed base, thus maximizing the number of people with which the user's files will be compatible. Furthermore, the user's training in a particular platform becomes more valuable as the size of the installed base of the platform increases. If the user must invest considerable effort in learning to use a computer platform, the user will probably choose to invest this effort in learning the format he or she believes will be most widely used.

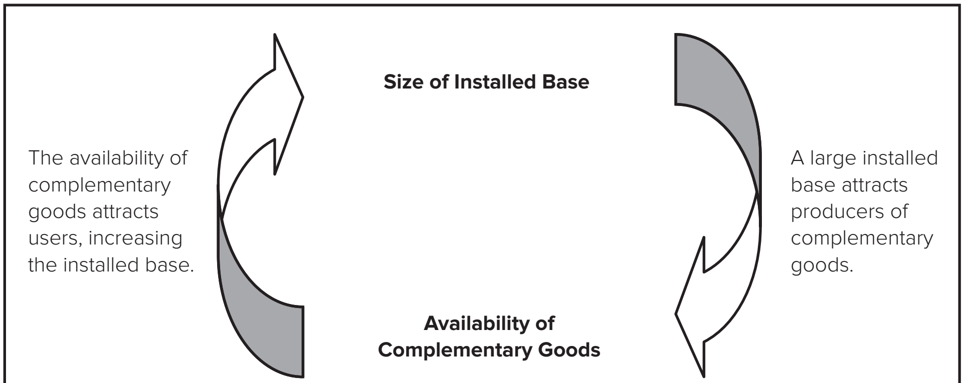
installed base The number of users of a particular good. For instance, the installed base of a particular video game console refers to the number of those consoles that are installed in homes worldwide.

Network externalities also arise when **complementary goods** are important. Many products are functional or desirable only when there is a set of complementary goods available for them (videotapes for VCRs, film for cameras, etc.). Some firms make both a good and its complements (e.g., Kodak produced both cameras and film), whereas others rely on other companies to provide complementary goods or services for their products (e.g., computer manufacturers often rely on other vendors to supply service and software to customers). Products that have a large installed base are likely to attract more developers of complementary goods. This is demonstrated in the Theory in Action about Microsoft: Once the Windows operating system had the largest installed base, most producers of complementary software applications chose to design their products to be optimized to work with Windows. Since the availability of complementary goods will influence users' choice among competing platforms, the availability of complementary goods influences the size of the installed base. A self-reinforcing cycle ensues (see Figure 4.2).

complementary goods Additional goods and services that enable or enhance the value of another good. For example, the value of a video game console is directly related to the availability of complementary goods such as video games, peripheral devices, and services such as online gaming.

The effect of this cycle is vividly demonstrated by Microsoft's dominance of the operating system market, and later the graphical user interface market, as discussed in the Theory in Action on the rise of Microsoft. Microsoft's early advantage in installed base led to an advantage in the availability of complementary goods. These network externality advantages enabled Windows to lock several would-be contenders such as Geoworks and NeXT (and, some would argue, Apple) out of the market.

FIGURE 4.2
The Self-Reinforcing Cycle of Installed Base and Availability of Complementary Goods



Theory in Action The Rise of Microsoft

From the early 1980s to the decade beginning in 2010, Microsoft's Windows controlled an overwhelming share of the personal computer operating system market. An operating system is the main program on a computer, which enables it to run other programs. Operating systems are responsible for recognizing the input from a keyboard, sending output to the display, tracking files and directories on the disk drives, and controlling peripheral devices. Because the operating system determines how other software applications must be designed, Microsoft's dominance in the operating system market made it extraordinarily powerful in the software industry. However, Microsoft's emergence as a software superpower was due largely to the unfolding of a unique set of circumstances. Had these events played out differently, Microsoft's dominance might have never been.

In 1980, the dominant operating system for personal computers was CP/M. CP/M was invented by Gary Kildall and marketed by Kildall's company, Digital Research. Kildall had been retained by Intel in 1972 to write software for Intel's 4004, the first true microprocessor in that it could be programmed to do custom calculations. Later that year, Intel began to sell the 8008 to designers who would use it as a computer, and Kildall was hired to write a programming language for the chip, called PL/M (Programming Language/Microcomputers).[a]

Then Memorex and Shugart began offering floppy disks (which IBM had invented) as a replacement for punch cards, and Kildall acquired one of these drives. However, no existing program would make the disk drive communicate with Intel's microprocessor, so he wrote a disk operating system that he called Control Program/Microprocessor (CP/M).[b] CP/M could be adapted to any computer based on Intel microprocessors.

Before 1980, IBM, the world's largest computer producer, had not been interested in developing a personal computer. IBM managers could not imagine the personal computer market ever amounting to more than a small niche of hobbyists. However, when businesses began adopting Apple computers to do basic accounting or word processing, IBM began to get nervous. IBM suddenly realized that the personal computer market might become a significant industry, and if it wanted to be a major player in that market it needed to act fast. IBM's managers did not believe they had time to develop their own microprocessor and operating system, so they based their personal computer on Intel microprocessors and planned to use Kildall's CP/M operating system. There are many stories of why Kildall did not sign with IBM. One story is that Kildall was out flying his plane when IBM came around, and though the IBM managers left their names with Kildall's wife, Dorothy McEwen, they did not state the nature of their business, and Kildall did not get back to them for some time. Another version of the story posits that Kildall was reluctant to become tied into any long-term contracts with the massive company, preferring to retain his independence. Yet a third version claims that Kildall was simply more interested in developing advanced technologies than in the strategic management of the resulting products. Whatever the reason, Kildall did not sign with IBM.

Pressed for time, IBM turned to Bill Gates, who was already supplying other software for the system, and asked if he could provide an operating system as well. Though Gates did not have an operating system at that time, he replied that he could supply one. Gates bought a 16-bit operating system (basically a clone of CP/M) from Seattle Computer Company, and reworked the software to match IBM's machines. The product was called Microsoft DOS. With DOS bundled on every IBM PC (which sold more than 250,000 units the first year), the product had an immediate and immense installed base. Furthermore, the companies that emerged to fulfill the unmet demand for IBM PCs with clones also adopted Microsoft DOS to ensure that their products were IBM PC compatible. Because it replicated CP/M, Microsoft DOS was compatible with the range of software that had been developed for the CP/M operating system. Furthermore, after it was bundled with the IBM PC, more software was developed for the operating system, creating an even wider availability of complementary goods. Microsoft DOS was soon entrenched as the industry standard, and Microsoft was the world's fastest-growing software company.

"We were able to get the technology out into the market early to develop a standard. We were effective in soliciting software vendors to write to that platform to solidify the standard," said B. J. Whalen, Microsoft product manager. "Once you get it going, it's a snowball effect. The more applications you have available for a

continued

concluded

platform, the more people will want to use that platform. And of course, the more people that want to use that platform, the more software vendors will want to write to that platform."

Later Microsoft would develop a graphical interface named Windows that closely replicated the user-friendly functionality of Apple computers. By bundling Windows with DOS, Microsoft was able to transition its base of DOS customers over to the Windows system. Microsoft also worked vigorously to ensure that compatible applications were developed for DOS and Windows, making applications itself and also encouraging third-party developers to support the platform. Microsoft was able to leverage its dominance with Windows into a major market share in many other software markets (e.g., word processing, spreadsheet programs, presentation programs) and influence over many aspects of the computer software and hardware industries. However, had Kildall signed with IBM, or had Compaq and other computer companies been unable to clone the IBM personal computer, the software industry might look very different today.

[a] P. Korzeniowski, "DOS: Still Thriving after All These Years," *Software Magazine* 10, no. 6 (1990), pp. 83–112.
[b] S. Veit, "What Ever Happened to . . . Gary Kildall?" *Computer Shopper* 14, no. 11 (1994), pp. 608–14.

Firms can also attempt to influence the selection of a dominant design by building coalitions around a preferred technology.[9] This is aptly illustrated in the opening case. While the preceding has emphasized the emergence of dominant designs through market forces, occasionally a dominant design is put in place through government regulation.

Government Regulation

In some industries, the consumer welfare benefits of having compatibility among technologies have prompted government regulation, and thus a legally induced adherence to a dominant design. This has often been the case for the utilities, telecommunications, and television industries, to name a few.[10] For example, in 1953, the U.S. Federal Communications Commission (FCC) approved the National Television Systems Committee (NTSC) color standard in television broadcasting to ensure that individuals with monochrome television sets would be able to receive the color television programs broadcast by networks (though they would see them in black and white). That standard was still in place in 2003. Similarly, in 1998, while a battle was being fought in the United States over wireless technology formats, the European Union (EU) adopted a single wireless telephone standard (the general standard for mobile communications, or GSM). By choosing a uniform standard, the EU could avoid the proliferation of incompatible standards and facilitate exchange both within and across national borders. Where government regulation imposes a single standard on an industry, the technology design embodied in that standard necessarily dominates the other technology options available to the industry. The consumer welfare impact of dominant designs is explored further in the Theory in Action section on Are Winner-Take-All Markets Good for Consumers?

The Result: Winner-Take-All Markets

All these forces can encourage the market toward natural monopolies. While some alternative platforms may survive by focusing on niche markets, the majority of the market may be dominated by a single (or few) design(s). A firm that is able to lock in its technology as the dominant design of a market usually earns huge rewards and may dominate the product category through several product generations. When a firm's technology is chosen as a dominant design, not only does the firm have the potential to earn near-monopoly rents in the short run, but the firm also is in a good position to shape the evolution of the industry, greatly influencing what future generations of products will look like. However, if the firm supports a technology that is not chosen as the dominant design, it may be forced to adopt the dominant technology, effectively forfeiting the capital, learning, and brand equity invested in its original technology. Even worse, a firm may find itself locked out of the market if it is unable to adopt the dominant technology. Such standards battles are high-stakes games—resulting in big winners and big losers.

path dependency
When end results depend greatly on the events that took place leading up to the outcome. It is often impossible to reproduce the results that occur in such a situation.

Increasing returns to adoption also imply that technology trajectories are characterized by **path dependency**, meaning that relatively small historical events may have a great impact on the final outcome. Though the technology's quality and technical advantage undoubtedly influence its fate, other factors, unrelated to the technical superiority or inferiority, may also play important roles.[11] For instance, timing may be crucial; early technology offerings may become so entrenched that subsequent technologies, even if considered to be technically superior, may be unable to gain a foothold in the market. How and by whom the technology is sponsored may also impact adoption. If, for example, a large and powerful firm aggressively sponsors a technology (perhaps even pressuring suppliers or distributors to support the technology), that technology may gain a controlling share of the market, locking out alternative technologies.

The influence of a dominant design can also extend beyond its own technology cycle. As the dominant design is adopted and refined, it influences the knowledge that is accumulated by producers and customers, and it shapes the problem-solving techniques used in the industry. Firms will tend to use and build on their existing knowledge base rather than enter unfamiliar areas.[12] This can result in a very "sticky" technological paradigm that directs future technological inquiry in the area.[13] Thus, a dominant design is likely to influence the nature of the technological discontinuity that will eventually replace it.

Such winner-take-all markets demonstrate very different competitive dynamics than markets in which many competitors can coexist relatively peacefully. These markets also require very different firm strategies to achieve success. Technologically superior products do not always win—the firms that win are usually the ones that know how to manage the multiple dimensions of value that shape design selection.

MULTIPLE DIMENSIONS OF VALUE

The value a new technology offers a customer is a composite of many different things. We first consider the value of the stand-alone technology, and then show how the stand-alone value of the technology combines with the value created by the size of the installed base and availability of complementary goods.[14] In industries characterized by increasing returns, this combination will influence which technology design rises to dominance.

A Technology's Stand-Alone Value

The value a new technology offers to customers can be driven by many different things, such as the functions it enables the customer to perform, its aesthetic qualities, and its ease of use. To help managers identify the different aspects of utility a new technology offers customers, W. Chan Kim and Renee Mauborgne developed a "Buyer Utility Map."[15] They argue that it is important to consider six different utility levers, as well as six stages of the buyer experience cycle, to understand a new technology's utility to a buyer.

The stages they identify are *purchase, delivery, use, supplements, maintenance,* and *disposal.* The six utility levers they consider are *customer productivity, simplicity, convenience, risk, fun and image,* and *environmental friendliness.* Creating a grid with stages and levers yields a 36-cell utility map (see Figure 4.3). Each cell provides an opportunity to offer a new value proposition to a customer.

A new technology might offer a change in value in a single cell or in a combination of cells. For example, when retailers establish an online ordering system, the primary new value proposition they are offering is greater *simplicity* in the *purchase* stage. On the other hand, as shown in Figure 4.3, the introduction of the Toyota Prius hybrid-electric vehicle offered customers greater productivity (in the form of gas savings), image benefits, and environmental friendliness in the customer's use, supplements, and maintenance stages, while providing the same simplicity and convenience of regular gasoline-only–powered vehicles.

Kim and Mauborgne's model is designed with an emphasis on consumer products, but their mapping principle can be easily adapted to emphasize industrial products or different aspects of buyer utility. For example, instead of having a single entry for customer productivity, the map could have rows for several dimensions of productivity such as speed, efficiency, scalability, and reliability. The map provides a guide for managers to consider multiple dimensions of technological value and multiple stages of the customer experience. Finally, the new benefits have to be considered with respect to the cost to the customer of obtaining or using the technology—it is the ratio of benefits to cost that determines value.

Network Externality Value

In industries characterized by network externalities, the value of a technological innovation to users will be a function not only of its stand-alone benefits and cost, but also of the value created by the size of its installed base and the availability of complementary goods (see Figure 4.4(a)).[16] Thus, the value to consumers of using the Windows operating system is due in part to the technology's stand-alone value (e.g., the ability of the operating system to make it easy for consumers to use the computer), the installed base of the operating system (and thus the number of computers with which the user can easily interact), and the availability of compatible software. Visualizing the value of technological innovations in this way makes it clear why even innovations that offer significant improvements in technological functionality often fail to displace existing technologies that are already widely adopted: Even if a new innovation has a significant advantage in functionality, its overall value may be significantly less than the incumbent standard. This situation is poignantly illustrated in the case of NeXT computers. In 1985, Steve

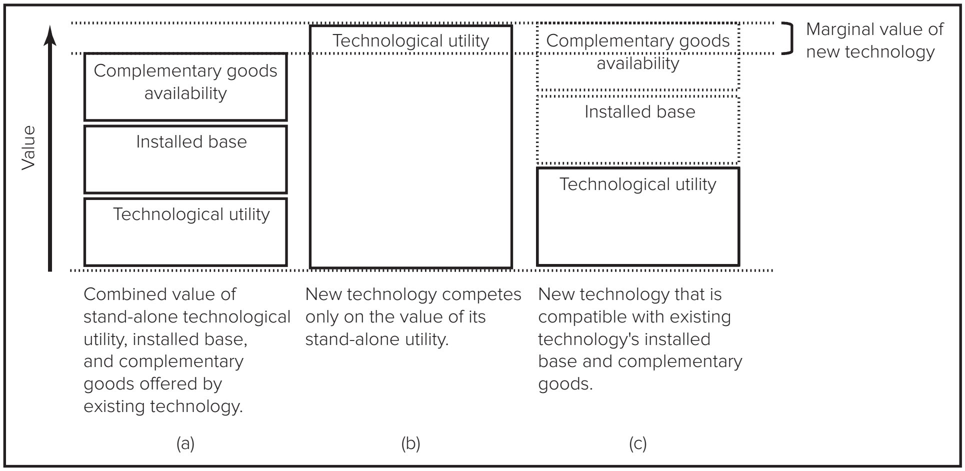
FIGURE 4.3
The Buyer Utility Map with Toyota Prius Example

Source: Adapted from *Harvard Business Review.* Exhibit from "Knowing a Winning Business Idea When You See One," by W. C. Kim and R. Mauborgne, September–October 2000.

	Purchase	Delivery	Use	Supplements	Maintenance	Disposal
Customer productivity	Price of Prius slightly higher than comparable nonhybrid models		Offers speed and power comparable to nonhybrid models	Can stop less often for gas, saving money and time		
Simplicity	Buyer may feel less able to assess value of vehicle		Operates like a regular combustion engine vehicle	Refuels like a regular combustion engine vehicle		Hybrids have larger batteries that would have to be recycled and disposed of at end of life
Convenience		Will be sold through traditional dealer channels	Does not have to be plugged into electrical outlet	Can purchase fuel at regular gas stations	Maintenance is similar to regular combustion engine vehicle	
Risk			Buyer might face a higher risk of product failure because it embodies a new technology		Buyer might have difficulty finding replacement parts because of new technology	Prius might be more difficult to resell or have lower resell value
Fun and image		Connotes image of environmental responsibility				
Environmental friendliness	Buyers feel they are helping support the development of more environmentally friendly cars		Emits lower levels of pollutants	Requires less use of fossil fuels		

Jobs and five senior managers of Apple Computer founded NeXT Incorporated. They unveiled their first computer in 1988. With a 25-MHz Motorola 68030 and 8 MB of RAM, the machine was significantly more powerful than most other personal computers available. It offered advanced graphics capability and even ran an object-oriented operating system (called NextStep) that was considered extremely advanced. However, the machine was not compatible with the IBM-compatible

FIGURE 4.4
Components of Value



personal computers (based on Intel's microprocessors and Microsoft's operating system) that had become the dominant standard. The machine thus would not run the vast majority of software applications on the market. A small contingent of early adopters bought the NeXT personal computers, but the general market rejected them because of a dire lack of software and uncertainty about the company's viability. The company discontinued its hardware line in 1993 and ceased development of NextStep in 1996.

A similar battle was playing out in 2015 between smartphone operating systems, though in this case there were two contenders who were more evenly matched: Apple's iOS and Google's Android. Both companies offered smartphone operating systems with intuitive, powerful, and aesthetically pleasing interfaces (technological utility). Both were aggressively building communities of applications providers that provided large ranges of interesting and/or useful applications (complementary goods). Both were also trying to build installed base through aggressive marketing and distribution. Market share estimates of the two systems varied widely based on the timing of the data announcements, the geographical scope considered, and the product scope considered, but in early 2015 it was clear that Apple and Google were in a head-to-head battle for dominance, whereas Rim's Blackberry and Microsoft's mobile operating systems were barely in the race (for more on this, see the section on the "Segment Zero" threat to Microsoft in Chapter Three).

As shown in Figure 4.4(b), it is not enough for a new technology's stand-alone utility to exceed that of the incumbent standard. The new technology must be able to offer greater overall value. For the new technology to compete on its stand-alone utility alone, that utility must be so great that it eclipses the combined value of an existing technology's stand-alone utility, its installed base, and its complementary goods.

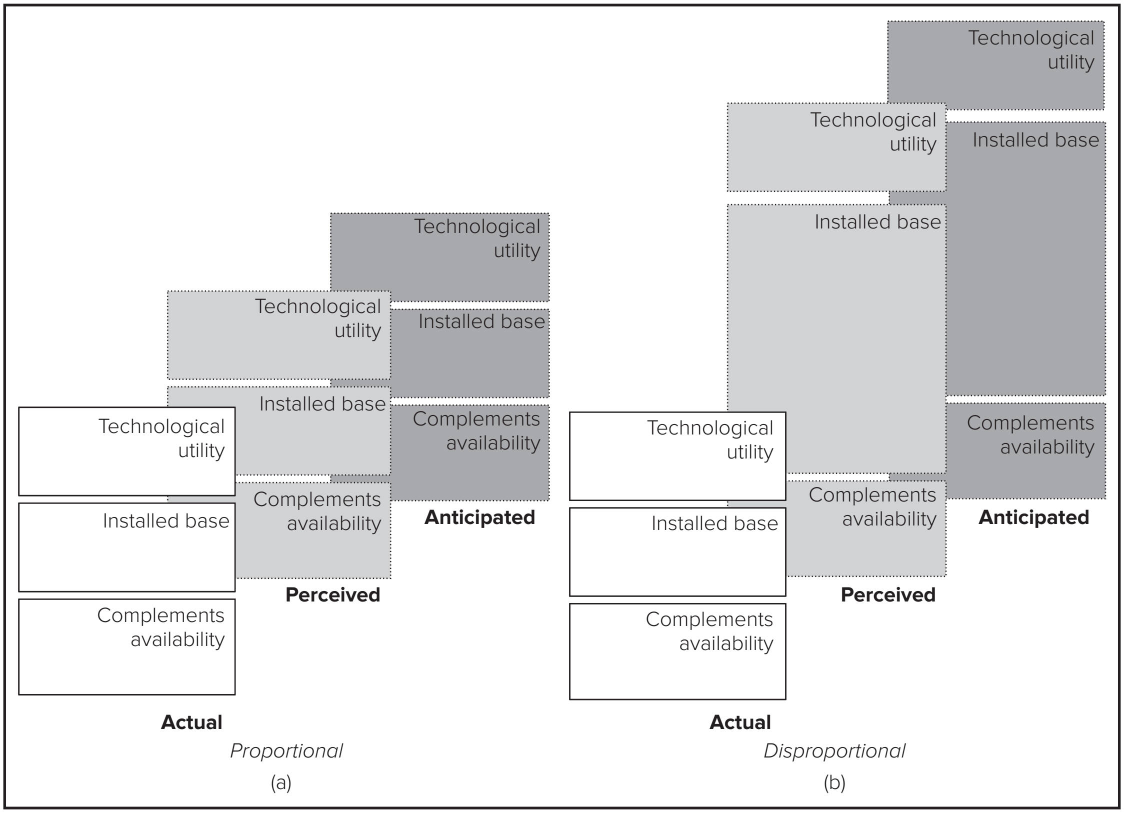
In some cases, the new technology may be made compatible with the existing technology's installed base and complementary goods as in Figure 4.4(c). In this case, a new technology with only a moderate functionality advantage may offer greater overall value to users. Sony and Philips employed this strategy with their high-definition audio format, Super Audio CD (SACD), a high-density multichannel audio format based on a revolutionary "scalable" bit-stream technology known as Direct Stream Digital (DSD). Anticipating that users would be reluctant to replace their existing compact disk players and compact disk music collections, Sony and Philips made the new Super Audio CD technology compatible with existing compact disk technology. The Super Audio CD players included a feature that enables them to play standard CDs, and the recorded Super Audio CDs included a CD audio layer in addition to the high-density layer, enabling them to be played on standard CD systems. Customers can thus take advantage of the new technology without giving up the value of their existing CD players and music libraries.

When users are comparing the value of a new technology to an existing technology, they are weighing a combination of objective information (e.g., actual technological benefits, actual information on installed base or complementary goods), subjective information (e.g., perceived technological benefits, perceived installed base or complementary goods), and expectations for the future (e.g., anticipated technological benefits, anticipated installed base and complementary goods). Thus, each of the primary value components described above also has corresponding perceived or anticipated value components (see Figure 4.5). In Figure 4.5(a), the perceived and anticipated value components map proportionately to their corresponding actual components. However, as depicted in Figure 4.5(b), this need not be the case. For instance, perceived installed base may greatly exceed actual installed base, or customers may expect that a technology will eventually have a much larger installed base than competitors and thus the value accrued from the technology's installed base is expected to grow much larger than it is currently.

Firms can take advantage of the fact that users rely on both objective and subjective information in assessing the combined value offered by a new technology. For example, even a technology with a small installed base can achieve a relatively large mind share through heavy advertising by its backers. Producers can also shape users' expectations of the future installed base and availability of complements through announcements of preorders, licensing agreements, and distribution arrangements. For example, when Sega and Nintendo were battling for dominance in the 16-bit video game console market, they went to great lengths to manage impressions of their installed base and market share, often to the point of deception. At the end of 1991, Nintendo claimed it had sold 2 million units of the Super Nintendo Entertainment System in the U.S. market. Sega disagreed, arguing that Nintendo had sold 1 million units at most. By May 1992, Nintendo was claiming a 60 percent share of the 16-bit market, and Sega was claiming a 63 percent share![17] Since perceived or expected installed base may drive subsequent adoptions, a large perceived or expected installed base can lead to a large actual installed base.

Such a tactic also underlies the use of "vaporware"—products that are not actually on the market and may not even exist but are advertised—by many software vendors. By building the impression among customers that a product is ubiquitous, firms can prompt rapid adoption of the product when it actually is available.

FIGURE 4.5
Actual, Perceived, and Expected Components of Value



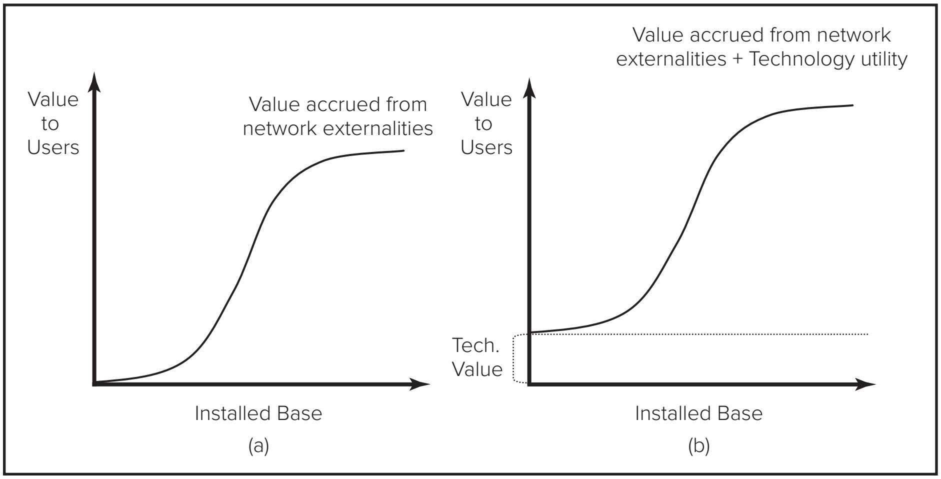
Vaporware may also buy a firm valuable time in bringing its product to market. If other vendors beat the firm to market and the firm fears that customers may select a dominant design before its offering is introduced, it can use vaporware to attempt to persuade customers to delay purchase until the firm's product is available. The video game console industry also provides an excellent example here. When Sega and Sony introduced their 32-bit video game consoles (the Saturn and PlayStation, respectively), Nintendo was still a long way from introducing its next-generation console. In an effort to forestall consumer purchases of 32-bit systems, Nintendo began aggressively promoting its development of a 64-bit system (originally named Project Reality) in 1994, though the product would not actually reach the market until September 1996. The project underwent so many delays that some industry observers dubbed it "Project Unreality."[18] Nintendo was successful in persuading many customers to wait for its Nintendo 64, and the system was ultimately relatively successful.

Nintendo, however, was never able to reclaim dominance over the video game industry. By the time the Nintendo 64 had gained significant momentum, Sony was developing its even more advanced PlayStation 2. Sony's experience in VCRs and compact disks had taught it to manage the multiple dimensions of value very well: Sony's PlayStation 2 offered more than double the processing power of the Nintendo 64, it was backward compatible (helping the PlayStation 2 tap the value of customers' existing PlayStation game libraries), and Sony sold it for a price that many speculated was less than the cost of manufacturing the console ($299). Sony also invested heavily to ensure that many game titles would be available at launch, and it used its distribution leverage and advertising budget to ensure the product would seem ubiquitous at its launch.

Competing for Design Dominance in Markets with Network Externalities

Graphs illustrate how differing technological utilities and network externality returns to installed base or market share impact the competition for design dominance. The following figures examine whether network externalities create pressure for a single dominant design versus a few dominant designs by considering the rate at which value increases with the size of the installed base, and how large of an installed base is necessary before most of the network externality benefits are achieved. As explained earlier, when an industry has network externalities, the value of a good to a user increases with the number of other users of the same or similar good. However, it is rare that the value goes up linearly—instead, the value is likely to increase in an s-shape as shown in Figure 4.6(a). Initially, the benefits may increase slowly. For example, whether a cell phone can reach 1 percent of the population or 5 percent is fairly insignificant—the reach of the phone service has to become much wider before the phone has much value. However, beyond some threshold level, the network externality returns begin to increase rapidly, until at some point, most of the benefits have been obtained and the rate of return decreases. Consider the example of operating systems at the beginning of the

FIGURE 4.6
Network Externality Returns to Market Share

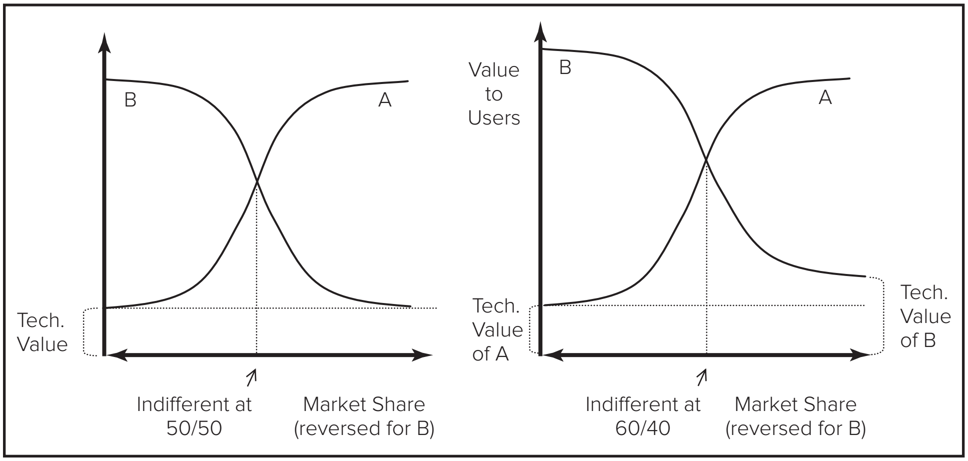


chapter: If an operating system has too small of an installed base, few software developers will write applications for it and thus it will be of little value to consumers. An increase from a 1 percent market share to a 2 percent market share makes little difference—developers are still unlikely to be attracted to the platform. Once the operating system exceeds some threshold level of adoption, however, it becomes worthwhile to develop software applications for it, and the value of the operating system begins to increase rapidly. Once the operating system achieves a large share of the market, the user has probably obtained most of the network externality value. There is likely to be a large range of quality software available for the operating system, and incremental increases in available software have less marginal impact on the value reaped by the customer.

Next we consider the stand-alone functionality of the technology. In Figure 4.6(b), a base level of technological utility has been added to the graph, which shifts the entire graph up. For example, an operating system that has an exceptionally easy-to-use interface makes the technology more valuable at any level of installed base. This becomes relevant later when two technologies that have different base levels of technological utility are considered.

When two technologies compete for dominance, customers will compare the overall value yielded (or expected) from each technology, as discussed in the previous section. In Figure 4.7, two technologies, A and B, each offer similar technological utility, and have similarly shaped network externality returns curves. To illustrate the competitive effects of two technologies competing for market share, the graphs in Figure 4.7 are drawn with market share on the horizontal axis instead of installed base. Furthermore, the curve for B is drawn with the market share dimension reversed so that we can compare the value offered by the two different technologies at different market share splits, that is, when A has a 20 percent market share, B has an 80 percent market share, and so on. This graph shows that at every point where A has less than 50 percent market share (and thus B has greater than 50 percent market share), B will yield greater overall value, making B more

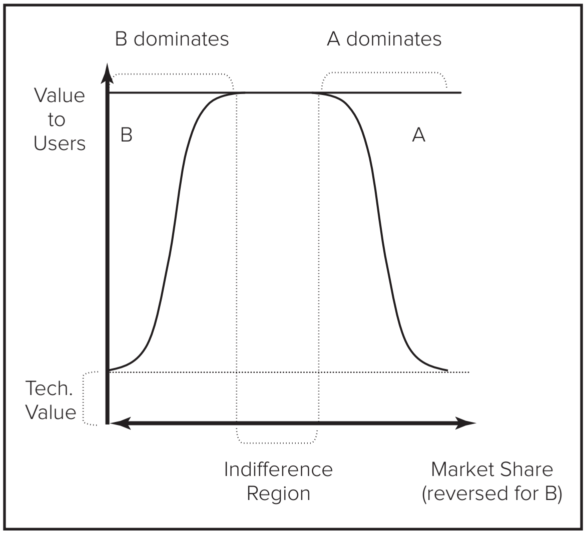
FIGURE 4.7
Network Externality Returns and Technological Utility: Competing Designs



attractive to customers. On the other hand, when A has greater than 50 percent market share (and B thus has less than 50 percent market share), A yields more overall value. When each technology has exactly 50 percent market share, they yield the same overall value and customers will be indifferent between them. However, if both technologies earn similar network externality returns to market share, but one technology offers greater stand-alone utility, the indifference point will be shifted in its favor. In the right-hand graph in Figure 4.7, technology B offers a greater level of stand-alone technological utility, shifting its overall value curve up. In this graph, technology A must have greater than 60 percent market share (and B must have less than 40 percent market share) for A to offer more overall value than B.

Another interesting scenario arises when customers attain their desired level of network externality benefits at lower levels of market share, depicted graphically in Figure 4.8. In this graph, the curves flatten out sooner, implying that the maximum amount of network externality value is obtained by customers at lower levels of market share. In this case, customers may face a relatively large indifference region within which neither technology clearly dominates. This may be the case with the video game console industry: While customers may experience some network externality benefits to a console having significant share (more game titles, more people to play against), those benefits might be achieved by a console without attaining a majority of the market. For example, even with Sony, Microsoft, and Nintendo splitting the game console market, there is still an abundance of game titles for all three consoles and a significant pool of people to play games against. Such markets may not experience great pressure to select a single dominant design; two or more platforms may successfully coexist.

FIGURE 4.8
Network Externality Value Is Fully Tapped at Minority Market Share Levels



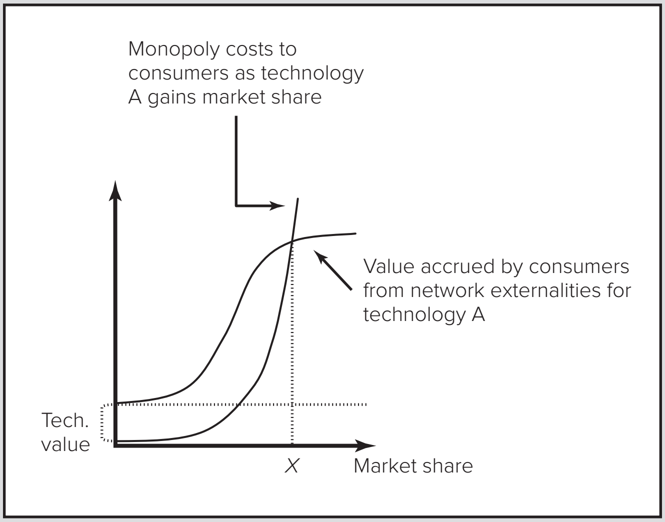
Theory in Action Are Winner-Take-All Markets Good for Consumers?

Traditionally, economics has emphasized the consumer welfare benefits of competitive markets; however, increasing returns make this a complicated issue. This is exemplified by the antitrust suits brought against Microsoft. While some analysts argued that Microsoft had clearly engaged in anticompetitive behavior and had damaged consumers in its quest to dominate the personal computer operating system market, others argued that Microsoft had behaved appropriately, and that its overwhelming share of the personal computer operating system market was good for consumers since it created greater compatibility among computers and more software applications. So how does a regulatory body decide when a firm has become too dominant? One way to think about this is to compare the value customers reap from network externalities at different levels of market share with the corresponding monopoly costs. Network externality returns refers to the value customers reap as a larger portion of the market adopts the same good (e.g., there is likely to be greater availability of complementary goods, more compatibility among users, and more revenues can be channeled into further developing the technology). Monopoly costs refer to the costs users bear as a larger portion of the market adopts the same good (e.g., a monopolist may charge higher prices, there may be less product variety, and innovation in alternative technologies may be stifled). Network externality returns to market share often exhibit the s-shape described in the previous section. Monopoly costs to market share, however, are often considered to be exponentially increasing. Plotting them on the same graph (as in Figure 4.9) reveals how network externality benefits and monopoly costs trade-off against each other.

In Figure 4.9, so long as technology A's market share remains less than *X*, the combination of technological utility and network externality benefits exceeds the monopoly costs, even if *X* represents a very large share of the market. However, as technology A's market share climbs beyond *X*, the monopoly costs now exceed the value of the technology utility and network externality benefits. A number of factors can shift where these two curves cross. If the technology utility for A were higher, the curves would cross at a point greater than *X*. If the network externality returns curve began to flatten at a lower market share (as was demonstrated earlier with the video game console industry), then the curves would cross at a market share less than *X*.

The steepness of the monopoly cost curve is largely a function of the firm's discretionary behavior. A firm can choose not to exploit its monopoly power, thus flattening

FIGURE 4.9
Network Externality Benefits and Monopoly Costs



continued

concluded

the monopoly costs curve. For instance, one of the most obvious assertions of monopoly power is typically exhibited in the price charged for a good. However, a firm can choose not to charge the maximum price that customers would be willing to pay for a good. For example, many people would argue that Microsoft does not charge the maximum price for its Windows operating system that the market would bear. However, a firm can also assert its monopoly power in more subtle ways, by controlling the evolution of the industry through selectively aiding some suppliers or complementors more than others, and many people would argue that in this respect, Microsoft has taken full advantage of its near-monopoly power.

MODULARITY AND PLATFORM COMPETITION

As noted earlier in the chapter, in many industries the value of a product system is strongly related to the number and quality of complements available for it. A smartphone operating system, for example, is only as good as the applications available, and a music streaming service is only as valuable as the number and quality and music titles it offers. In markets like these, rather than having a single firm produce both the platform (e.g., the smartphone operating system) and all of the complements (e.g., applications for the smartphone), industry players will instead often use *modularity* to create a *platform ecosystem* where many different firms contribute to the product system.[19]

Modularity

modularity
The degree to which a system's components can be separated and recombined.

To understand **modularity**, consider a product like a computer: A computer is a bundle of components that includes a microprocessor, a monitor, a keyboard, and more. Some of these components can be bought separately and assembled by the user, and some are typically bought preassembled. Each of those components can also be thought of as a bundle. The monitor, for example, includes circuitry, plastic housing, power input, and so on. Most of these are not sold separately to the end user, but they might be mixed and matched by a manufacturer to achieve different end products.

Products may be made increasingly modular both by expanding the range of compatible components (increasing the range of possible product configurations), and by uncoupling integrated functions within components (making the product modular at a finer level) (see Figure 4.10). For example, a smartphone manufacturer might originally offer only proprietary phones where they have produced both the hardware and software, and integrate them tightly into a single product configuration (Figure 4.10, panel A). However, greater market demand for flexibility might induce manufacturers to begin offering phones with a few different product configurations, each composed of the firm's own components. Should customers prefer to be able to combine phones with peripheral devices or applications developed by other producers, smartphone makers may eventually "open" their systems up, creating a standard by which other developers can create products that are compatible with the phones (Figure 4.10, panel B). Smartphone makers may even decide to uncouple their operating systems from the hardware so that consumers can use the operating system on devices made by other manufacturers (Figure 4.10, panel C). In each of these stages, the product has become increasingly modular.

FIGURE 4.10
Modularity and Platform Ecosystems

A. Traditional integrated product bundle:
- Provider tries to meet buyers needs itself
- No customization, no external compatibility

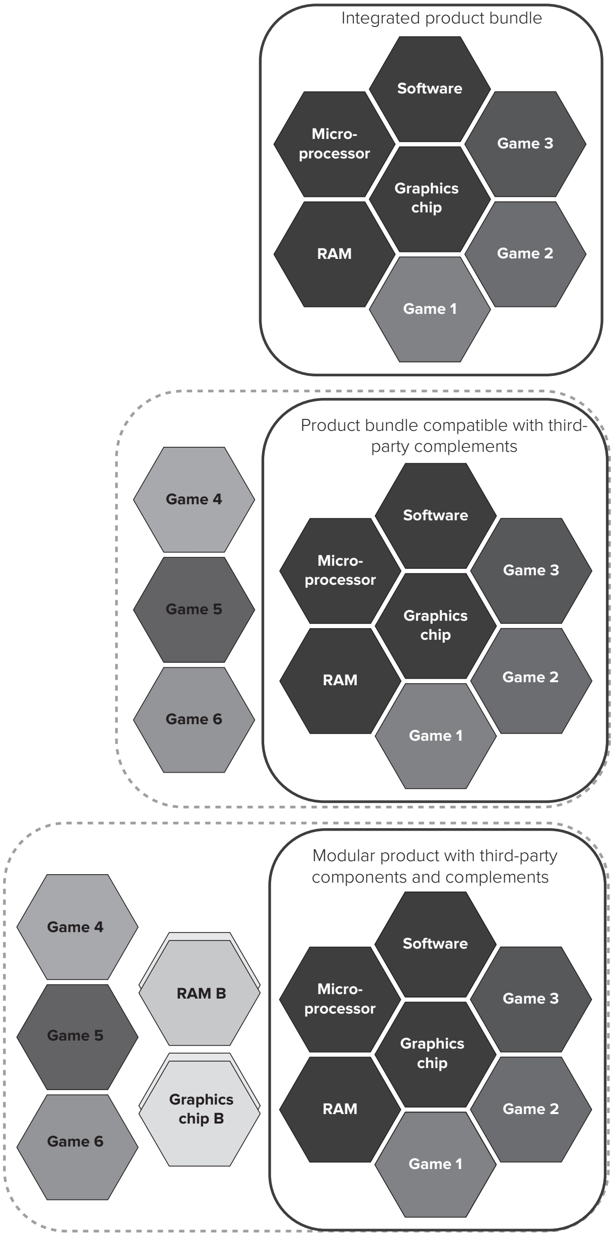
Example: Nokia E90 Communicator

B. Product bundle with third-party complements:
- Compatibility with third-party choices expands options for customers

Example: Apple iPhone

C. Product bundle with third-party components and complements:
- Customer has even greater range of configuration choices

Example: Android device



The majority of products are modular at some level. For example, many stereos, shelving systems, and bicycles allow the customer to mix and match components from different vendors. There are a great many more products that are modular in that the customer can choose components but not assemble the end product themselves. For example, when you buy a car, you can often choose an engine size, upholstery options, automatic steering or transmission, stereo system, tires, roof racks, security systems, and so on, but the automaker assembles the configuration for you. Even aircraft are offered this way: though Boeing and Airbus manufacture airframes, they do not manufacture engines. Engines are produced by companies such as GE, Pratt Whitney, or Rolls Royce. The engines are generally designed to be used in a few different aircraft models. Similarly, the aircraft is not committed to a single engine, though they may designate a "launch engine" as a preferred choice. The aircraft customer typically makes the final decision on which engine will be used in the aircraft.

Tightly integrated (i.e., nonmodular) product systems and modular systems have different kinds of advantages. A tightly integrated product system might have components that are customized to work together, which may enable a level of performance that more standardized components cannot achieve. The producer of a tightly integrated system also has more control over the end product, which can enable them to better monitor quality and reliability. For years this was the reason Steve Jobs gave for not making Apple computers as modular as Windows-based computers—he believed that by controlling all of the components and most of the software, Apple computers could achieve greater functionality and be more reliable. An integrated product may also be more attractive to a customer that does not want to choose or assemble components themselves.

Modular products, on the other hand, often offer more choices over function, design, scale, and other features, enabling the customer to choose a product system that more closely suits their needs and preferences. Second, because components are reused in different combinations, this can achieve product variety while still allowing scale economies in manufacturing the individual components. This is known as *economies of substitution.*[20]

platform ecosystems
Ecosystem is a contraction of *ecological* and *system*, and refers to a system where elements share some form of mutual dependence. A platform in this context is a stable core that mediates the relationship between a range of components, complements, and end users. Thus *platform ecosystem* refers to a system of mutually dependent entities mediated by a stable core.

Modularity becomes increasingly valuable in a product system when there are (a) diverse technological options available to be recombined, and (b) heterogeneous customer preferences.[21] For example, there is a very wide range of applications available for smartphones, and customers are very heterogeneous in the applications they want on their smartphones. This increases the value of being able to pick and choose your own customized mix of applications that go on your smartphone. This example also reveals how pressure for modularity can lead to platform ecosystems, as discussed below.

Platform Ecosystems

The word *ecosystem* comes from biology and is a contraction of *ecological system.* This means that when we use the term *ecosystem*, we are usually referring to entities that have some degree of mutual dependence. In a **platform ecosystem**, some core part of a product (such as a video game console or music streaming service) mediates the relationship between a wide range of other components or complements (such as video games or music titles) and prospective end users.[22]

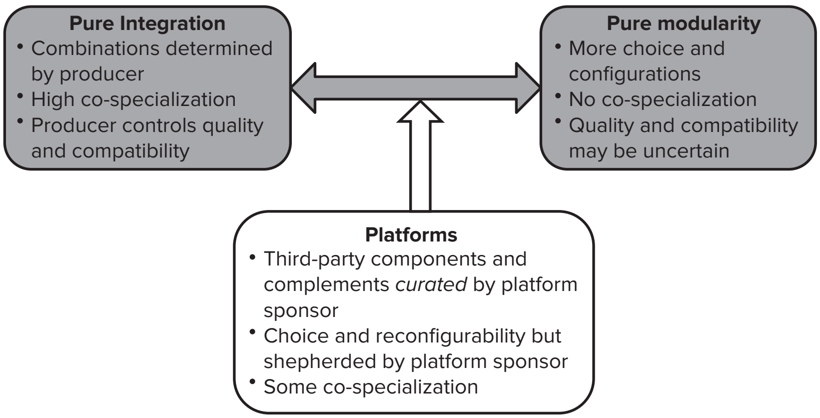
A platform's boundaries can be well-defined with a stable set of members dedicated wholly to that platform, or they can be amorphous and changing, with members entering and exiting freely, and participating in multiple platforms simultaneously. For example, consider the difference between the television/movie streaming services HBO on Demand and Amazon Prime. HBO on Demand exists to serve only HBO content up to consumers. The shows available are tightly controlled, and there is limited entry and exit of show producers. The Amazon Prime ecosystem is much more open. In fact, just about any content producer—including individual independent filmmakers—can make their content available on Amazon Prime.

Because it is the overall appeal of the ecosystem that attracts end users to the platform, the success of individual members depends, at least in part, upon the success of other members of the ecosystem—even those with which they may be simultaneously competing. Furthermore, in many platforms there are switching costs that make it difficult or costly to change ecosystems. Platforms and their complements often make investments in co-specialization or sign exclusivity agreements that bind them into stickier, longer-term relationships than the market contracts used in typical reseller arrangements. A video game that has been made for the Microsoft Xbox, for example, cannot be played on a PlayStation console unless a new version of it is made (and the game producer may have signed a contract with Microsoft that prohibits this).

A platform ecosystem is thus characterized by relationships that are neither as independent as arms-length market contracts nor as dependent as those within a hierarchical organization. It is, in essence, a hybrid organizational form.[23] It strikes a compromise between the loose coupling of a purely modular system and the tight coupling of a traditional integrated product. It enables customers to mix and match some components and complements, while still enabling some co-specialization and curation of the complements and components available for the system (see Figure 4.11).

Once we understand that platforms are like a compromise between pure modularity and pure integration, it becomes easier to understand when platforms will be desirable in a market. First, platforms will be more valuable than a tightly integrated product when (a) customers are diverse and want more choices than a single firm can provide, (b) when third-party options are diverse and high quality, (c) when compatibility with third-party products can be made seamless without integration, and/or (d) when the platform sponsor is powerful enough that it can retain control over quality and the overall product architecture without producing the complements itself. Looking at it from the other direction, platforms will be more valuable than a purely modular system when (a) complements are nonroutine purchases with uncertainty (and thus the

FIGURE 4.11
Platforms as a Compromise between Pure Modularity and Pure Integration



customer prefers to have some shepherding by the platform sponsor), (b) when some integration between the platform and its complements provides performance advantages, and/or (c) when important components of the ecosystem require subsidization (e.g., the market is unlikely to provide all the complements the end customer needs at adequate quality or value).

Video game systems are an iconic example of platform ecosystems. Consoles need to launch with high-quality games. Since it is difficult to induce game developers to make games for a console that has not yet been widely adopted, most game console producers must produce games themselves (or subsidize their production) to ensure that high-quality games are available when the console launches. On the other hand, end users want more games than just those produced by the console producer, so console producers like Microsoft, Sony, and Nintendo also license third-party developers to produce games for their consoles. They carefully screen the licensed games for quality and compatibility, and they may require the game developers to sign exclusivity agreements or to customize the games for the console. The console maker may also manage the end users' awareness and perception of the games in the ecosystem by giving "Best" of awards to particular games, by bundling particular games with the console at point of purchase, or by featuring particular games in its marketing. These strategies enable the console producer to actively manage the overall value created by its ecosystem.[24]

Summary of Chapter

1. Many technologies demonstrate increasing returns to adoption, meaning that the more they are adopted, the more valuable they become.
2. One primary source of increasing returns is learning-curve effects. The more a technology is produced and used, the better understood and developed it becomes, leading to improved performance and reduced costs.
3. Another key factor creating increasing returns is network externality effects. Network externality effects arise when the value of a good to a user increases with the size of the installed base. This can be due to a number of reasons, such as need for compatibility or the availability of complementary goods.
4. In some industries, the consumer welfare benefits of having a single standard have prompted government regulation, such as the European Union's mandate to use the GSM cellular phone standard.
5. Increasing returns can lead to winner-take-all markets where one or a few companies capture nearly all the market share.
6. The value of a technology to buyers is multidimensional. The stand-alone value of a technology can include many factors (productivity, simplicity, etc.) and the technology's cost. In increasing returns industries, the value will also be significantly affected by the technology's installed base and availability of complementary goods.
7. Customers weigh a combination of objective and subjective information. Thus, a customer's perceptions and expectations of a technology can be as important as (or more important than) the actual value offered by the technology.

8. Firms can try to manage customers' perceptions and expectations through advertising and public announcements of preorders, distribution agreements, and so on.
9. The combination of network externality returns to market share and technological utility will influence at what level of market share one technology will dominate another. For some industries, the full network externality benefits are attained at a minority market share level; in these industries, multiple designs are likely to coexist.
10. In markets where customers have heterogeneous preferences and there are many potential technological options available, firms might use modularity to enable customers to mix and match components, producing a wider array of end products.
11. Platform ecosystems are an example of modularity in action. A stable core platform (e.g., a smartphone operating system) may mediate the relationship between many complements producers (e.g., applications developers) and end users.

Discussion Questions

1. What are some of the sources of increasing returns to adoption?
2. What are some examples of industries not mentioned in the chapter that demonstrate increasing returns to adoption?
3. What are some of the ways a firm can try to increase the overall value of its technology and its likelihood of becoming the dominant design?
4. What determines whether an industry is likely to have one or a few dominant designs?
5. Are dominant designs good for consumers, competitors, complementors, and suppliers?

Suggested Further Reading

Classics

Arthur, W. B., *Increasing Returns and Path Dependency in the Economy* (Ann Arbor, MI: University of Michigan Press, 1994).

Gawer, A., and M. A. Cusumano, *Platform Leadership* (Boston: Harvard Business School Publishing, 2002).

Katz, M., and C. Shapiro, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy* 94 (1986):822–41.

Schilling, M. A., "Technological Leapfrogging: Lessons from the U.S. Video Game Console Industry," *California Management Review* 45, no. 3 (2003):6–32.

Shapiro, C., and Varian, H. *Information Rules: A Strategic Guide to the Network Economy* (Boston: Harvard Business Review Press, 1999).

Schilling, M. A., "Towards a General Modular Systems Theory and Its Application to Inter-Firm Product Modularity," *Academy of Management Review* 25 (2000):312–34.

Recent Work

Adner, R., and R. Kapoor, "Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations," *Strategic Management Journal* 31 (2010):306–33.

Boudreau, K. J., "Open Platform Strategies and Innovation: Granting Access vs. Devolving Control," *Management Science* 56 (2010):1849–72.

Boudreau, K. J., and L. B. Jeppesen, "Unpaid Crowd Complementors: The Platform Network Effect Mirage," *Strategic Management Journal* 36, no. 12 (2015):1761–77.

Jacobides, M., C. Cennamo, and A. Gawer, "Towards a Theory of Ecosystems," *Strategic Management Journal* 39 (May 2018):2255–76.

Sterman, J. D., R. Henderson, E. D. Beinhocker, and L. I. Newman, "Getting Big Too Fast: Strategic Dynamics with Increasing Returns and Bounded Rationality," *Management Science* 53 (2007):683–96.

Seamans, R., and F. Zhu, "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," *Management Science* 60 (2013):476–93.

Endnotes

1. W. B. Arthur, *Increasing Returns and Path Dependency in the Economy* (Ann Arbor, MI: University of Michigan Press, 1994).
2. For examples of this, see M. Lapre, A. S. Mukherjee, and L. N. Van Wassenhove, "Behind the Learning Curve: Linking Learning Activities to Waste Reduction," *Management Science* 46 (2000), pp. 597–611; F. K. Levy, "Adaptation in the Production Process," *Management Science* 11 (1965), pp. B136–54; and L. E. Yelle, "The Learning Curve: Historical Review and Comprehensive Survey," *Decision Sciences* 10 (1979), pp. 302–28.
3. L. Argote, *Organizational Learning: Creating, Retaining and Transferring Knowledge* (Boston: Kluwer Academic Publishers, 1999); N. W. Hatch and D. C. Mowery, "Process Innovation and Learning by Doing in Semiconductor Manufacturing," *Management Science* 44, no. 11 (1998), pp. 1461–77; and M. A. Schilling, P. Vidal, R. Ployhart, and A. Marangoni, "Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve," *Management Science* 49 (2003), pp. 39–56.
4. For examples, see L. Argote, "Group and Organizational Learning Curves: Individual, System and Environmental Components," *British Journal of Social Psychology* 32 (1993), pp. 31–52; Argote, *Organizational Learning;* N. Baloff, "Extensions of the Learning Curve—Some Empirical Results," *Operations Research Quarterly* 22, no. 4 (1971), pp. 329–40; E. D. Darr, L. Argote, and D. Epple, "The Acquisition, Transfer and Depreciation of Knowledge in Service Organizations: Productivity in Franchises," *Management Science* 41 (1995), pp. 1750–62; L. Greenberg, "Why the Mine Injury Picture Is Out of Focus," *Mining Engineering* 23 (1971), pp. 51–53; Hatch and Mowery, "Process Innovation and Learning by Doing in Semiconductor Manufacturing"; A. Mukherjee, M. Lapre, and L. Wassenhove, "Knowledge Driven Quality Improvement," *Management Science* 44 (1998), pp. S35–S49; and Yelle, "The Learning Curve."
5. Argote, *Organizational Learning*.
6. J. Dutton and A. Thomas, "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review* 9 (1984), pp. 235–47; Levy, "Adaptation in the Production Process"; and Mukherjee, Lapre, and Wassenhove, "Knowledge Driven Quality Improvement."
7. W. M. Cohen and D. A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly,* March 1990, pp. 128–52.
8. M. Katz and C. Shapiro, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy* 94 (1986), pp. 822–41; M. Schilling, "Technological Lock Out: An Integrative Model of the Economic and Strategic Factors Driving Technology Success and Failure," *Academy of Management Review* 23 (1998), pp. 267–84; and M. Thum, "Network

Externalities, Technological Progress, and the Competition of Market Contracts," *International Journal of Industrial Organization* 12 (1994), pp. 269–89.

9. J. Wade, "Dynamics of Organizational Communities and Technological Bandwagons: An Empirical Investigation of Community Evolution in the Microprocessor Market," *Strategic Management Journal* 16 (1995), pp. 111–34.
10. Schilling, "Technological Lock Out"; and F. F. Suarez, "Battles for Technological Dominance: An Integrative Framework," *Research Policy* 33, pp. 271–86.
11. W. B. Arthur, "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *The Economic Journal,* March 1989, pp. 116–31; R. W. England, "Three Reasons for Investing Now in Fossil Fuel Conservation: Technological Lock-In, Institutional Inertia, and Oil Wars," *Journal of Economic Issues,* September 1994, pp. 755–76; and Katz and Shapiro, "Technology Adoption in the Presence of Network Externalities."
12. G. Dosi, "Sources, Procedures, and Microeconomic Effects of Innovation," *Journal of Economic Literature* 26 (1988), p. 1130.
13. Ibid., pp. 1120–71.
14. M. A. Schilling, "Technological Leapfrogging: Lessons from the U.S. Video Game Console Industry," *California Management Review* 45, no. 3 (2003), pp. 6–32; and Suarez, "Battles for Technological Dominance."
15. W. C. Kim and R. Mauborgne, "Knowing a Winning Business Idea When You See One," *Harvard Business Review,* September–October 2000, pp. 129–38.
16. Schilling, "Technological Leapfrogging."
17. A. Brandenberger, 1995b. "Power Play (B): Sega in 16-Bit Video Games," Harvard Business School case no. 9-795-103.
18. A. Brandenberger, 1995c. "Power Play (C): 3DO in 32-Bit Video Games," Harvard Business School case no. 9-795-104; and Schilling, "Technological Leapfrogging."
19. Ceccagnoli, M., C. Forman, P. Huang, D. J. Wu, "Cocreation of Values in a Platform Ecosystem: The Case of Enterprise Software." *MIS Quart* 36, no. 1 (2012):263–90; Gawer, A., "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework." *Research Policy*, 43, no. 7 (2014):1239–49; Jacobides, M. G., Knudsen, T., Augier, M., "Benefiting from Innovation: Value Creation, Value Appropriation and the Role of Industry Architectures." *Research Policy*, 35, no. 8(2006):1200–21; Rietveld, J., M. A. Schilling, and C. Bellavitis, "Platform Strategy: Managing Ecosystem Value Through Selective Promotion of Complements." (working paper, New York University, 2018); Schilling, M. A., "Towards a General Modular Systems Theory and Its Application to Inter-Firm Product Modularity." *Academy of Management Review*, 25 (2000):312–34.
20. Garud, R., and A. Kumaraswamy, "Technological and Organizational Designs for Realizing Economies of Substitution." *Strategic Management Journal* 16 (1995):93–109.
21. Schilling, M. A., "Towards a General Modular Systems Theory and Its Application to Inter-Firm Product Modularity." *Academy of Management Review* 25 (2000):312–34.
22. Rietveld, J., M. A. Schilling, and C. Bellavitis, "Platform Strategy: Managing Ecosystem Value Through Selective Promotion of Complements." (working paper, New York University, 2018).
23. Ibid.
24. Ibid.