Harvard Business Review

Finance And Investing

Increasing Returns and the New World of Business

by W. Brian Arthur

From the Magazine (July-August 1996)

Our understanding of how markets and businesses operate was passed down to us more than a century ago by a handful of European economists—Alfred Marshall in England and a few of his contemporaries on the continent. It is an understanding based squarely upon the assumption of diminishing returns: products or companies that get ahead in a market eventually run into limitations, so that a predictable equilibrium of prices and market shares is reached. The theory was roughly valid for the bulkprocessing, smokestack economy of Marshall's day. And it still thrives in today's economics textbooks. But steadily and continuously in this century, Western economies have undergone a transformation from bulk-material manufacturing to design and use of technology—from processing of resources to processing of information, from application of raw energy to application of ideas. As this shift has occurred, the underlying mechanisms that determine economic behavior have shifted from ones of diminishing to ones of *increasing* returns.

Increasing returns are the tendency for that which is ahead to get further ahead, for that which loses advantage to lose further advantage. They are mechanisms of positive feedback that

operate—within markets, businesses, and industries—to reinforce that which gains success or aggravate that which suffers loss. Increasing returns generate not equilibrium but instability: If a product or a company or a technology—one of many competing in a market—gets ahead by chance or clever strategy, increasing returns can magnify this advantage, and the product or company or technology can go on to lock in the market. More than causing products to become standards, increasing returns cause businesses to work differently, and they stand many of our notions of how business operates on their head.

Mechanisms of increasing returns exist alongside those of diminishing returns in all industries. But roughly speaking, diminishing returns hold sway in the traditional part of the economy—the processing industries. Increasing returns reign in the newer part—the knowledge-based industries. Modern economies have therefore bifurcated into two interrelated worlds of business corresponding to the two types of returns. The two worlds have different economics. They differ in behavior, style, and culture. They call for different management techniques, strategies, and codes of government regulation.

They call for different understandings.

Alfred Marshall's World

Let's go back to beginnings—to the diminishing-returns view of Alfred Marshall and his contemporaries. Marshall's world of the 1880s and 1890s was one of bulk production: of metal ores, aniline dyes, pig iron, coal, lumber, heavy chemicals, soybeans, coffee —commodities heavy on resources, light on know-how. In that world it was reasonable to suppose, for example, that if a coffee plantation expanded production it would ultimately be driven to use land less suitable for coffee. In other words, it would run into diminishing returns. So if coffee plantations competed, each one would expand until it ran into limitations in the form of rising costs or diminishing profits. The market would be shared by many plantations, and a market price would be established at a

predictable level—depending on tastes for coffee and the availability of suitable farmland. Planters would produce coffee so long as doing so was profitable, but because the price would be squeezed down to the average cost of production, no one would be able to make a killing. Marshall said such a market was in perfect competition, and the economic world he envisaged fitted beautifully with the Victorian values of his time. It was at equilibrium and therefore orderly, predictable and therefore amenable to scientific analysis, stable and therefore safe, slow to change and therefore continuous. Not too rushed, not too profitable. In a word, mannerly. In a word, genteel.

With a few changes, Marshall's world lives on a century later within that part of the modern economy still devoted to bulk processing: of grains, livestock, heavy chemicals, metals and ores, foodstuffs, retail goods—the part where operations are largely repetitive day to day or week to week. Product differentiation and brand names now mean that a few companies rather than many compete in a given market. But typically, if these companies try to expand, they run into some limitation: in numbers of consumers who prefer their brand, in regional demand, in access to raw materials. So no company can corner the market. And because such products are normally substitutable for one another, something like a standard price emerges. Margins are thin and nobody makes a killing. This isn't exactly Marshall's perfect competition, but it approximates it.

The Increasing-Returns World

What would happen if Marshall's diminishing returns were reversed so that there were *increasing* returns? If products that got ahead thereby got further ahead, how would markets work?

Let's look at the market for operating systems for personal computers in the early 1980s when CP/M, DOS, and Apple's Macintosh systems were competing. Operating systems show increasing returns: if one system gets ahead, it attracts further software developers and hardware manufacturers to adopt it,

which helps it get further ahead. CP/M was first in the market and by 1979 was well established. The Mac arrived later, but it was wonderfully easy to use. DOS was born when Microsoft locked up a deal in 1980 to supply an operating system for the IBM PC. For a year or two, it was by no means clear which system would prevail. The new IBM PC—DOS's platform—was a kludge. But the growing base of DOS/IBM users encouraged software developers such as Lotus to write for DOS. DOS's prevalence—and the IBM PC's bred further prevalence, and eventually the DOS/IBM combination came to dominate a considerable portion of the market. That history is now well known. But notice several things: It was not predictable in advance (before the IBM deal) which system would come to dominate. Once DOS/IBM got ahead, it locked in the market because it did not pay for users to switch. The dominant system was not the best: DOS was derided by computer professionals. And once DOS locked in the market, its sponsor, Microsoft, was able to spread its costs over a large base of users. The company enjoyed killer margins.

These properties, then, have become the hallmarks of increasing returns: market instability (the market tilts to favor a product that gets ahead), multiple potential outcomes (under different events in history, different operating systems could have won), unpredictability, the ability to lock in a market, the possible predominance of an inferior product, and fat profits for the winner. They surprised me when I first perceived them in the late 1970s. They were also repulsive to economists brought up on the order, predictability, and optimality of Marshall's world. Glimpsing some of these properties in 1939, English economist John Hicks warned that admitting increasing returns would lead to "the wreckage of the greater part of economic theory." But Hicks had it wrong: the theory of increasing returns does not destroy the standard theory—it complements it. Hicks felt repugnance not just because of unsavory properties but also because in his day no mathematical apparatus existed to analyze increasing-returns markets. That situation has now changed. Using sophisticated techniques from qualitative dynamics and

probability theory, I and others have developed methods to analyze increasing-returns markets. The theory of increasing returns is new, but it already is well established. And it renders such markets amenable to economic understanding.

In 1939, English economist John Hicks warned that admitting increasing returns would lead to "the wreckage of the greater part of economic theory." But Hicks had it wrong.

In the early days of my work on increasing returns, I was told they were an anomaly. Like some exotic particle in physics, they might exist in theory but would be rare in practice. And if they did exist, they would last for only a few seconds before being arbitraged away. But by the mid-1980s, I realized increasing returns were neither rare nor ephemeral. In fact, a major part of the economy was subject to increasing returns—high technology.

Why should this be so? There are several reasons:

Up-front Costs.

High-tech products—pharmaceuticals, computer hardware and software, aircraft and missiles, telecommunications equipment, bioengineered drugs, and suchlike—are by definition complicated to design and to deliver to the marketplace. They are heavy on know-how and light on resources. Hence they typically have R&D costs that are large relative to their unit production costs. The first disk of Windows to go out the door cost Microsoft \$50 million; the second and subsequent disks cost \$3. Unit costs fall as sales increase.

Network Effects.

Many high-tech products need to be compatible with a network of

users. So if much downloadable software on the Internet will soon appear as programs written in Sun Microsystems' Java language, users will need Java on their computers to run them. Java has competitors. But the more it gains prevalence, the more likely it will emerge as a standard.

Customer Groove-in.

High-tech products are typically difficult to use. They require training. Once users invest in this training—say, the maintenance and piloting of Airbus passenger aircraft—they merely need to update these skills for subsequent versions of the product. As more market is captured, it becomes easier to capture future markets.

In high-tech markets, such mechanisms ensure that products that gain market advantage stand to gain further advantage, making these markets unstable and subject to lock-in. Of course, lock-in is not forever. Technology comes in waves, and a lock-in such as DOS's can last only as long as a particular wave lasts.

Some products—like the IBM PC—start in the increasing-returns world but later in their life cycle become virtual commodities that belong to Marshall's processing world.

So we can usefully think of two economic regimes or worlds: a bulk-production world yielding products that essentially are congealed resources with a little knowledge and operating according to Marshall's principles of diminishing returns, and a knowledge-based part of the economy yielding products that essentially are congealed knowledge with a little resources and operating under increasing returns. The two worlds are not neatly split. Hewlett-Packard, for example, designs knowledge-based

devices in Palo Alto, California, and manufactures them in bulk in places like Corvallis, Oregon, or Greeley, Colorado. Most high-tech companies have both knowledge-based operations and bulk-processing operations. But because the rules of the game differ for each, companies often separate them—as Hewlett-Packard does. Conversely, manufacturing companies have operations such as logistics, branding, marketing, and distribution, which belong largely to the knowledge world. And some products—like the IBM PC—start in the increasing-returns world but later in their life cycle become virtual commodities that belong to Marshall's processing world.

The Halls of Production and the Casino of Technology

Because the two worlds of business—processing bulk goods and crafting knowledge into products—differ in their underlying economics, it follows that they differ in their character of competition and their culture of management. It is a mistake to think that what works in one world is appropriate for the other.

There is much talk these days about a new management style that involves flat hierarchies, mission orientation, flexibility in strategy, market positioning, reinvention, restructuring, reengineering, repositioning, reorganization, and re-everything else. Are these new insights or are they fads? Are they appropriate for all organizations? Why are we seeing this new management style?

Let us look at the two cultures of competition. In bulk processing, a set of standard prices typically emerges. Production tends to be repetitive—much the same from day to day or even from year to year. Competing therefore means keeping product flowing, trying to improve quality, getting costs down. There is an art to this sort of management, one widely discussed in the literature. It favors an environment free of surprises or glitches—an environment characterized by control and planning. Such an environment requires not just people to carry out production but also people to plan and control it. So it favors a hierarchy of bosses and workers.

Because bulk processing is repetitive, it allows constant improvement, constant optimization. And so, Marshall's world tends to be one that favors hierarchy, planning, and controls. Above all, it is a world of optimization.

Competition is different in knowledge-based industries because the economics are different. If knowledge-based companies are competing in winner-take-most markets, then managing becomes redefined as a series of quests for the next technological winner the next cash cow. The goal becomes the search for the Next Big Thing. In this milieu, management becomes not production oriented but mission oriented. Hierarchies flatten not because democracy is suddenly bestowed on the workforce or because computers can cut out much of middle management. They flatten because, to be effective, the deliverers of the next-thing-for-thecompany need to be organized like commando units in small teams that report directly to the CEO or to the board. Such people need free rein. The company's future survival depends upon them. So they—and the commando teams that report to them in turn—will be treated not as employees but as equals in the business of the company's success. Hierarchy dissipates and dissolves.

Does this mean that hierarchy should disappear in meatpacking, steel production, or the navy? Contrary to recent management evangelizing, a style that is called for in Silicon Valley will not necessarily be appropriate in the processing world. An aircraft's safe arrival depends on the captain, not on the flight attendants. The cabin crew can usefully be "empowered" and treated as human beings. This approach is wise and proper. But forever there will be a distinction—a hierarchy—between cockpit and cabin crews.

In fact, the style in the diminishing-returns Halls of Production is much like that of a sophisticated modern factory: the goal is to keep high-quality product flowing at low cost. There is little need to watch the market every day, and when things are going

smoothly the tempo can be leisurely. By contrast, the style of competition in the increasing-returns arena is more like gambling. Not poker, where the game is static and the players vie for a succession of pots. It is casino gambling, where part of the game is to choose which games to play, as well as playing them with skill. We can imagine the top figures in high tech—the Gateses and Gerstners and Groves of their industries—as milling in a large casino. Over at this table, a game is starting called multimedia. Over at that one, a game called Web services. In the corner is electronic banking. There are many such tables. You sit at one. How much to play? you ask. Three billion, the croupier replies. Who'll be playing? We won't know until they show up. What are the rules? Those'll emerge as the game unfolds. What are my odds of winning? We can't say. Do you still want to play?

High technology, pursued at this level, is not for the timid.

In fact, the art of playing the tables in the Casino of Technology is primarily a psychological one. What counts to some degree—but only to some degree—is technical expertise, deep pockets, will, and courage. Above all, the rewards go to the players who are first to make sense of the new games looming out of the technological fog, to see their shape, to cognize them. Bill Gates is not so much a wizard of technology as a wizard of precognition, of discerning the shape of the next game.

Adaptation means watching for the next wave and positioning the company to take advantage of it. Adaptation is what drives increasing-returns businesses, not optimization.

We can now begin to see that the new style of management is not a fad. The knowledge-based part of the economy demands flat hierarchies, mission orientation, above all a sense of direction.

Not five-year plans. We can also fathom the mystery of what I've alluded to as re-everything. Much of this "re-everything" predilection—in the bulk-processing world—is a fancy label for streamlining, computerizing, downsizing. However, in the increasing-returns world, especially in high tech, re-everything has become necessary because every time the quest changes, the company needs to change. It needs to reinvent its purpose, its goals, its way of doing things. In short, it needs to adapt. And adaptation never stops. In fact, in the increasing-returns environment I've just sketched, standard optimization makes little sense. You cannot optimize in the casino of increasingreturns games. You can be smart. You can be cunning. You can position. You can observe. But when the games themselves are not even fully defined, you cannot optimize. What you can do is adapt. Adaptation, in the proactive sense, means watching for the next wave that is coming, figuring out what shape it will take, and positioning the company to take advantage of it. Adaptation is what drives increasing-returns businesses, not optimization.

Playing the High-Tech Tables

Suppose you are a player in the knowledge-industry casino, in this increasing-returns world. What can you do to capitalize on the increasing returns at your disposal? How can you use them to capture markets? What strategic issues do you need to think about? In the processing world, strategy typically hinges upon capitalizing on core competencies, pricing competitively, getting costs down, bringing quality up. These are important also in the knowledge-based world, but so, too, are other strategies that make use of the special economics of positive feedbacks.

Two maxims are widely accepted in knowledge-based markets: it pays to hit the market first, and it pays to have superb technology. These maxims are true but do not guarantee success. Prodigy was first into the on-line services market but was passive in building its subscriber base to take advantage of increasing returns. As a result, it has fallen from its leading position and currently lags the other services. As for technology, Steve Jobs's NeXT workstation

was superb. But it was launched into a market already dominated by Sun Microsystems and Hewlett-Packard. It failed. A new product often has to be two or three times better in some dimension—price, speed, convenience—to dislodge a locked-in rival. So in knowledge-based markets, entering first with a fine product can yield advantage. But as strategy, this is still too passive. What is needed is *active* management of increasing returns.

One active strategy is to discount heavily initially to build up an installed base. Netscape handed out its Internet browser for free and won 70% of its market. Now it can profit from spin-off software and applications. Although such discounting is effective—and widely understood—it is not always implemented. Companies often err by pricing high initially to recoup expensive R&D costs. Yet even smart discounting to seed the market is ineffective unless the resulting installed base is exploited later. America Online built up a lead of more than 4.5 million subscribers by giving away free services. But because of the Internet's dominance, it is not yet clear whether it can transform this huge base into later profits.

Let's get a bit more sophisticated. Technological products do not stand alone. They depend on the existence of other products and other technologies. The Internet's World Wide Web operates within a grouping of businesses that include browsers, online news, E-mail, network retailing, and financial services. Pharmaceuticals exist within a network of physicians, testing labs, hospitals, and HMOs. Laser printers are part of a grouping of products that include computers, publishing software, scanners, and photo-input devices. Unlike products of the processing world, such as soybeans or rolled steel, technological products exist within local groupings of products that support and enhance them. They exist in mini-ecologies.

This interdependence has deep implications for strategy. When, in the mid-1980s, Novell introduced its network-operating

system, NetWare, as a way of connecting personal computers in local networks, Novell made sure that NetWare was technically superior to its rivals. It also heavily discounted NetWare to build an installed base. But these tactics were not enough. Novell recognized that NetWare's success depended on attracting software applications to run on NetWare—which was a part of the ecology outside the company's control. So it set up incentives for software developers to write for NetWare rather than for its rivals. The software writers did just that. And by building NetWare's success, they ensured their own. Novell managed these crossproduct positive feedbacks actively to lock in its market. It went on to profit hugely from upgrades, spin-offs, and applications of its own.

In the Case of Microsoft...

What should be legal in this powerful and as yet unregulated world of increasing returns? What constitutes fair ...

 \searrow

Another strategy that uses ecologies is linking and leveraging. This means transferring a user base built up upon one node of the ecology (one product) to neighboring nodes, or products. The strategy is very much like that in the game Go: you surround neighboring markets one by one, lever your user base onto them, and take them over—all the time enhancing your position in the industry. Microsoft levered its 60-million-person user base in DOS onto Windows, then onto Windows 95, and then onto Microsoft Network by offering inexpensive upgrades and by bundling applications. The strategy has been challenged legally. But it recognizes that positive feedbacks apply across markets as well as within markets.

In fact, if technological ecologies are now the basic units for strategy in the knowledge-based world, players compete not by locking in a product on their own but by building webs—loose alliances of companies organized around a mini-ecology—that amplify positive feedbacks to the base technology. Apple, in closing its Macintosh system to outsiders in the 1980s, opted not to create such a web. It believed that with its superior technology, it could hold its increasing-returns market to itself. Apple indeed dominates its Mac-based ecology. But this ecology is now only 8% of the personal computer business. IBM erred in the other direction. By passively allowing other companies to join its PC web as clones, IBM achieved a huge user base and locked in the market. But the company itself wound up with a small share of the spoils. The key in web building is active management of the cross-company mutual feedbacks. This means making a careful choice of partners to build upon. It also means that, rather than attempting to take over all products in the ecology, dominant players in a web should allow dependent players to lock in their dependent products by piggybacking on the web's success. By thus ceding some of the profits, the dominant players ensure that all participants remain committed to the alliance.

Important also to strategy in knowledge-based markets is psychological positioning. Under increasing returns, rivals will back off in a market not only if it is locked in but if they *believe* it will be locked in by someone else. Hence we see psychological jockeying in the form of preannouncements, feints, threatened alliances, technological preening, touted future partnerships, parades of vaporware (announced products that don't yet exist). This posturing and puffing acts much the way similar behavior does in a primate colony: it discourages competitors from taking on a potentially dominant rival. No moves need be made in this strategy of premarket facedown. It is purely a matter of psychology.

What if you hold a losing hand? Sometimes it pays to hold on for residual revenue. Sometimes a fix can be provided by updated

technology, fresh alliances, or product changes. But usually under heavy lock-in, these tactics do not work. The alternatives are then slow death or graceful exit—relinquishing the field to concentrate on positioning for the next technology wave. Exit may not mean quitting the business entirely. America Online, Compuserve, Prodigy, and Microsoft Network have all ceded dominance of the on-line computer networking market to the Internet. But instead of exiting, they are steadily becoming adjuncts of the Net, supplying content services such as financial quotations or games and entertainment. They have lost the main game. But they will likely continue in a side game with its own competition for dominance within the Net's ecology.

Above all, strategy in the knowledge world requires CEOs to recognize that a different kind of economics is at work. CEOs need to understand which positive and negative feedback mechanisms are at play in the market ecologies in which they compete. Often there are several such mechanisms—interbraided, operating over different time frames, each needing to be understood, observed, and actively managed.

What About Service Industries?

So far, I've talked mainly about high tech. Where do service industries such as insurance, restaurants, and banking fit in? Which world do they belong to? The question is tricky. It would appear that such industries belong to the diminishing-returns, processing part of the economy because often there are regional limits to the demand for a given service, most services do consist of "processing" clients, and services are low-tech.

The truth is that network or user-base effects often operate in services. Certainly, retail franchises exist because of increasing returns. The more McDonald's restaurants or Motel 6 franchises are out there geographically, the better they are known. Such businesses are patronized not just for their quality but also because people want to know exactly what to expect. So the more prevalent they are, the more prevalent they can become.

Similarly, the larger a bank's or insurance company's customer base, the more it can spread its fixed costs of headquarters staff, real estate, and computer operations. These industries, too, are subject to mild increasing returns.

So we can say more accurately that service industries are a hybrid. From day to day, they act like bulk-processing industries. But over the long term, increasing returns will dominate—even though their destabilizing effects are not as pronounced as in high tech. The U.S. airline business, for example, processes passengers day to day. So it seemed in 1981 that deregulation should enhance competition, as it normally does under diminishing returns. But over the long term, airlines in fact experience a positive feedback: under the hub-and-spoke system, once an airline gets into trouble, it cannot work the feeder system for its routes properly, its fleet ages, it starts a downward spiral, and it loses further routes. The result of deregulation over the long term has been a steady decline in large carriers, from 15 airlines in 1981 to approximately 6 at present. Some routes have become virtual monopolies, with resulting higher fares. None of this was intended. But it should have been predicted—given increasing returns.

In fact, the increasing-returns character of service industries is steadily strengthening. One of the marks of our time is that in services everything is going software—everything that is information based. So operations that were once handled by people—designing fancy financial instruments or automobiles or fashion goods, processing insurance claims, supplying and inventorying in retail, conducting paralegal searches for case precedents—are increasingly being handled by software. As this reengineering of services plays out, centralized software facilities come to the fore. Service providers become hitched into software networks, regional limitations weaken, and user-base network effects kick in.

This phenomenon can have two consequences. First, where the

local character of service remains important, it can preserve a large number of service companies but clustered round a dominant software provider—like the large numbers of small, independent law firms tied in to the dominant computer-search network, Lexis-Nexis. Or physicians tied in to an HMO. Second, where locality is unimportant, network effects can transform competition toward the winner-take-most character we see in high tech. For example, when Internet-based retail banking arrives, regional demand limitations will vanish. Each virtual bank will gain in advantage as its network increases. Barring regulation, consumer banking will then become a contest among a few large banking networks. It will become an increasing-returns business.

Services belong to both the processing and the increasing-returns world. But their center of gravity is crossing over to the latter.

Thoughts for Managers

Where does all this leave us? At the beginning of this century, industrial economies were based largely on the bulk processing of resources. At the close of the century, they are based on the processing of resources *and* on the processing of knowledge. Economies have bifurcated into two worlds—intertwined, overlapping, and different. These two worlds operate under different economic principles. Marshall's world is characterized by planning, control, and hierarchy. It is a world of materials, of processing, of optimization. The increasing-returns world is characterized by observation, positioning, flattened organizations, missions, teams, and cunning. It is a world of psychology, of cognition, of adaptation.

Many managers have some intuitive grasp of this new increasingreturns world. Few understand it thoroughly. Here are some questions managers need to ask themselves when they operate in knowledge-based markets:

Do I understand the feedbacks in my market?

In the processing world, understanding markets means understanding consumers' needs, distribution channels, and rivals' products. In the knowledge world, success requires a thorough understanding of the self-negating and self-reinforcing feedbacks in the market—the diminishing-and increasing-returns mechanisms. These feedbacks are interwoven and operate at different levels in the market and over different time frames.

Which ecologies am I in?

Technologies exist not alone but in an interlinked web, or ecology. It is important to understand the ecologies a company's products belong to. Success or failure is often decided not just by the company but also by the success or failure of the web it belongs to. Active management of such a web can be an important magnifier of increasing returns.

Do I have the resources to play?

Playing one of the increasing-returns games in the Casino of Technology requires several things: excellent technology, the ability to hit the market at the right time, deep pockets, strategic pricing, and a willingness to sacrifice current profits for future advantage. All this is a matter not just of resources but also of courage, resolution, will. And part of that resolution, that courage, is also the decisiveness to leave the market when increasing returns are moving against one. Hanging on to a losing position that is being further eroded by positive feedbacks requires throwing reinforcements into a battle already lost. Better to exit with financial dignity.

What games are coming next?

Technology comes in successive waves. Those who have lost out on this wave can position for the next. Conversely, those who have made a killing on this cycle should not become complacent. The ability to profit under increasing returns is only as good as the ability to see what's coming in the next cycle and to position

