HARVARD | BUSINESS | SCHOOL

9-611-082
MAY 13, 2011

WILLY SHIH

Scale Effects, Network Effects, and Investment Strategy

The transitions of many products from a centralized architecture where one firm supplies the complete solution into a modular architecture where specialists focus on specific vertical slices of the value network can have interesting consequences. What once might have been an integrated offering for a customer now takes the form of a platform and "complements," that somehow need to be coordinated, either deliberately or by the marketplace. Firms that specialize in one slice are confronted with developing new business models and more complex supply-demand dynamics. More pressing is the need for managers to now look beyond the sales of just their own products and into their value networks in order to understand how the market dynamics of the complements might influence the market performance of their own slice.

There are many products that fit into this category. The IBM Personal Computer (PC) is the textbook example. Architected as a modular system, the cloning of the IBM PC[1] gave rise to an industry of vertical specialists who focused on only a specific segment: microprocessors (Intel), hard disk drives (Seagate, Western Digital, IBM[2]), memory chips (Samsung, Elpida,[3] Micron), motherboards,[4] add-in cards, software, and system houses (Compaq,[5] HP, Dell, IBM[6]) that specialized in assembly and sales. Both Intel and Microsoft, once they recognized the transformed landscape, aggressively managed their complements and the scope of their own business models to foster the growth of the market and produce extraordinary returns for themselves.

Many other products and services fit into this modular system description. Videocassette recorders, telecommunications equipment, many transportation networks, Apple's iPhone and iPod, and payment cards (Visa, Mastercard, American Express) all represent a slice of a value chain with complementary pieces that are in the hands of others. To become large and successful, the platform owners need the complementary pieces to have the motivation, economic or otherwise, to grow in concert.

Entrepreneurs who are trying to start businesses that fit this model are often faced with the task of building infrastructure or an installed base of users before they can earn a return on their investments. Terms like "increasing returns to scale" and "network effects" are frequently used to justify big up-front investments to build out infrastructure well in advance of revenues. Success stories like Amazon.com in online retailing and eBay's online auction are a siren song to entrepreneurs and investors. But many of the same types of scale and scope arguments were used to justify investments in WebVan, an online grocery business that spent nearly $1 billion to build a network of warehouses and to purchase a fleet of delivery trucks before going bust in 2001. Other

---

Professor Willy Shih prepared this note as the basis for class discussion. Professors Andrei Hagiu and David Yoffie provided valuable input.

notable business model disasters include Furniture.com, whose product shipping costs reached 50% of revenues, Pets.com, which sold pet food and accessories over the Internet,[7] the online toy retailer eToys.com,[8] and Boo.com, an online clothing retailer.[9]

If we want to build a sustainable business model for this type of business, it is worth thinking through how these businesses scale. The purpose of this note is to expose the thinking behind scale economies and network effects, so that we have can use better theory for building and judging business models.[10] We start with some definitions, and then we look at the causality underlying scale economies and network effects so that we can try to carefully describe the circumstances under which businesses might exhibit one or the other or both. Finally we will look at *impatient for profit, patient for growth* and see how that theory works in conjunction to prescribe investment strategy.

Scale Economies

Economies of scale are cost decreases that result from expanding production. If the cost per unit of output rises more slowly than the costs of inputs in the same proportions, there are economies of scale (for example, if output doubles while the total cost of inputs less than doubles). If the costs rise more quickly than the cost of proportionate inputs, there are diseconomies of scale. The simplest scale economies arise when there are high setup costs and relatively low run costs per unit of output so that spreading setup over larger run volumes improves efficiency of labor and asset utilization.

Economies of scale are frequently found in industries that require large capital expenditures on plant and equipment, or the establishment of a large infrastructure prior to the ability to begin providing service. High fixed costs get apportioned across the entire product volumes, so larger production volumes mean a smaller per unit allocation. Economies of scale also can be found in circumstances where there is a low or no fixed cost, but the marginal cost of production declines with increasing volumes. A simple example of this circumstance arises from bulk buying of raw materials that are inputs to production.

Economies of scale are often a justification to "get big fast." Competing on scale is more prevalent in some industries like natural resources, many kinds of manufacturing including automotive, aerospace, petrochemicals, and semiconductors, and in some service industries like data center operations or ocean container shipping where labor content is a relatively small fraction of the total value add. Eventually diminishing returns sets in when the cost of complexity of managing the large scale rises faster than the savings. The other risk with "get big fast" is that large scale operations are vulnerable to capacity underutilization.

Direct and Indirect Network Effects

The rise of information and communications technologies, in particular the Internet, has popularized phenomena known as network effects. While scale effects operate primarily on the cost side, network effects operate on the "willingness to pay" side. The former is mostly inside the boundaries of what the firm controls, the latter are described as *externalities*.

When there is a shift from vertically integrated towards modular – and the basis of competition shifts from performance to configurability and flexibility – industry structures often shift as well to favor firms who specialize in *platforms,* which serve as the foundation of a product or service offering, and the associated firms who produce *complements*. Often such platforms generate little value on their own; they only become valuable to a customer as part of a *system* that incorporates these

complementary components. A PC is worth little without software, but as a platform that is used in conjunction with that software it can serve a wide range of markets. Music players rely on pre-recorded music, and a credit card is useless without a transaction network behind it. We refer to the complementary component like music and movies, or even the provision of a service as "software" for the purposes of this discussion. The economics of these system markets have attracted considerable interest because of the wide range of business models that purport to be based on the effects.[11]

The challenge for the manager is to coordinate complementing products provided by the value network. There are two generic types of coordination that are valuable to system builders. The first is when the consumer connects the product to a network as part of its usage model. Telephones and fax machines are the prime examples. Such networks are valuable precisely because other consumers acquire similar components and join the same network. Owning a single fax machine in isolation is worth little, but when many people purchase fax machines and connect then to the network, each individual fax machine becomes worth more to its respective owner. This phenomenon – the increase in value of belonging to a network that is a consequence of other users joining and enlarging the network – is called a *network effect*. Benefits to buyers correlate positively with the total number of people who buy compatible products.

The other type of coordination is when consumers must choose a durable hardware platform. Often buyers will spread out the purchase of components of the system over time. They care about the buying decisions of others because of the effect that it has on the cost and availability of the complementing products. Systems that are expected to be popular will be perceived to have more widely available or less expensive components, and for those very reasons will become more popular. This is known as an *indirect network effect*. Expectations about price, performance, quality, availability and other attributes can strongly influence whether consumers want to "buy in" to the system. This positive feedback means there is the potential to reach a "tipping point" where one system begins to achieve market dominance. Consumers perceive that once a system choice is made, switching later on can be expensive. Choosing between an Apple Mac and a Microsoft Windows based PC is a classic systems choice decision, as are the choices between satellite, terrestrial, and Internet radio, or between the Sony PlayStation, Microsoft Xbox, and Nintendo game console systems. Once buyers realized that the VHS videocassette format had a much wider availability of movies to rent than the rival Betamax, the VCR market reached a tipping point. After that, Betamax went into a rapid decline. PC platforms, audio-video equipment, and many types of capital equipment coupled with repair services fit this model,[12] though tipping points are reached far less frequently than many people assume.

Indirect network effects arise in systems where there are economies of specialization in "horizontal layers" of a system (*e.g.* operating system for PCs), made possible by technology advances and modularization, and when consumer demand for variety favors competition based on flexibility and configurability. As such they are "demand side" economies since they relate to demand and customers' willingness to pay, and not the supply side.

There is a special case of the two component system in which both components are produced exclusively by the same company. If the components are used in a fixed proportion, say one part hardware to one part software, this is not different than as if the product were one. But if they are used in variable proportions, as in a razor that will consume many blades over its lifetime or an inkjet printer that uses many ink cartridges, firms have the opportunity to price the components separately in a way unrelated to their actual cost to manufacture, providing the firm with two pricing instruments instead of one. How a firm distributes the pricing, and the timing of customer purchases

of the complementing component are cornerstones of this "razor and blade" business model. Many product companies aspire to this kind of "consumables" business model because of the economic similarity to monopoly pricing. Economists describe companies in this set of circumstances as an "integrated monopolist."

Multi-sided Platforms

Another lens for looking at indirect network effects frames the product as a *multi-sided platform,* which coordinates demand between different groups of customers who need each other in some way; they operate as market intermediaries who reduce search or transaction costs.[13] Most electronic entertainment devices, PCs, mobile phones and smart phones like the iPhone, eBay, Internet web portals and even shopping malls and newspapers fall into this category. **Exhibit 1** lists some more examples of multisided platforms.

The business model of a multi-sided platform is a combination of pricing and strategic decisions that affect each side. The pricing structure a firm chooses, including how much it will charge each side, and how it allocates costs to one side or the other can be a critical determinant of the success or failure of the model. In some cases firms treat one side of a multi-sided platform as a profit center and the other side as a loss leader. Sometimes the business model on one side of a platform will take longer to reach a planned state than the other side, but in the steady state the aggregate of all the models will ultimately determine overall profitability. So firms have been known to "dig a big hole" on one side or the other waiting to make all the models mesh. Firms may even choose to lose money on one side beyond the start-up phase and into the steady state simply because they can enhance the overall value by promoting the indirect network effect.

When developing a business model for a multi-sided platform, firms usually have to address several issues:

- The order to get the sides on board. The platform owner, say a game console manufacturer, faces a chicken-and-egg problem: with no game console available, there will be no software, but with no software they will sell no game consoles. Payment cards like Visa and American Express face this problem. There would be no demand among households for cards if you could not use them at lots of merchants, and no demand by merchants if lots of customers did not have them.

- The platform design – which features to offer each side so as to reduce search or transaction costs or both.

- Governance – who gets access to what side, and what each side is allowed to do on the platform.

- Pricing structure – many platforms have to cope with demand that may not be balanced across sides. Firms often end up with models heavily skewed to one side of the market; the margin is much less on one side than the other. For example, dating clubs have to match the numbers of men and women, and they sometimes will let women in for free while charging men. The total volume of transactions and the profit earned on a platform depends on the total charged to all of the participating parties as well as how the price is decomposed among parties. An example of the pricing for a payment card network is shown in **Exhibit 2**.

Examples of Business Models in Businesses with Network Effects

It is easy to imagine indirect network effects in many businesses. Some are strong, but many are weak. Assessing the strength is important to understanding whether the market has the potential to tip. The differences in outcomes can be dramatic.

Television, newspapers, magazines. The business model of free over-the-air broadcast television, and to a large extent newspapers and magazines, has been to provide free or inexpensive access to one side of the platform (viewers, readers) while charging the other side (advertisers). There have been some instances of an indirect network effect, as when a large program audience creates buzz and draws additional viewers, but by and large traditional media benefits from pure scale effects while network effects are weak.

Internet portals have adopted the free content model with a vengeance, providing free content, free e-mail, free search and other services while charging advertisers for banner advertising, sponsored search, keyword search, and referral fees. While some Internet portals purport to exhibit network effects, the ones where the effect is most clear are online communities like Facebook in which the service becomes more valuable to users when overall there are more members.

The Microsoft Windows software platform. Microsoft makes money selling the operating system to users (and achieves economies of scale by selling in large quantities to system builders). To a first approximation, the company does not appear to either make or lose substantial amounts of money on applications developers. It invests heavily in development tools. The company makes substantial profit on its Microsoft Office applications suite, and both Windows and Office see positive indirect network effects.

The console video game market. The business model in this market is for platform companies like Sony or Nintendo to price consoles to consumers at a loss, and to charge game developers for software development kits and a per copy royalty on games sold for that platform. Interestingly, Microsoft has entered this market with its Xbox platform but tries to leverage the existing development tools for its Windows platform. It collects royalties from game developers. This is a sharp contrast to the company's strategy in the PC market.

Adobe Portable Document Format (PDF). Adobe introduced PDF in 1993 as a universal way of representing two-dimensional documents that was independent of the application that created it or the hardware or software that might be used to view it. Its Acrobat software was intended to be a suite that could create and read PDF files. Early adoption was slow, but then Adobe made its Acrobat Reader freely available at no cost, and adoption started taking off. By the time Adobe submitted Acrobat Reader to the International Standards Organization in 2008 as a proposed international standard, it had become a *de facto* standard. Adobe did not reap much profit from PDF initially, but starting around 1998-99 it realized that it could reposition it as a platform with PDF as the file format and Acrobat Reader as the rendering platform, developing and selling other applications that would complement the pair.[14] The company tried to focus on higher value-add applications that it could monetize.[15]

Emergent Strategies: a Key to Success

All of the entrepreneurs who ended up with dot com disasters surely had business plans that they sold to venture capitalists. What theories were they using? Did they assume they would benefit from scale or network effects without investing the effort to understand the circumstances under

which theory did predict such benefits? What theories were the VCs using? How might one avoid such disasters?

The complexity of multi-sided platforms means that many assumptions about the business models on respective sides need to not only be correct, but they must work in aggregate net of all cross subsidies. The urgency of some managers to get both sides on board often causes them to misprice one side or skew a model to one side early on. This is what led many firms to offer "free" products or services to one side, in the hopes that they could ultimately balance all the models on the platform later as the other side(s) developed. Skewing *per se* is not necessarily a problem, but skewing towards the wrong side may be difficult to rectify later. Many models have failed when entrepreneurs have sacrificed profits in an early stage, hoping to raise prices later. In other cases managers have overestimated the amount of a network effect, or assumed that they could exclusively appropriate the returns.

The complexity of intersecting models makes it risky to apply deliberate strategies like "get big fast" when the supporting data is probably missing or suspect. Looking at the wreckage of some of the dot com disasters, one can speculate that perceived scale or network effect benefits that accrued from size made them impatient for growth and perhaps a little too patient for profit.

A look at some of the most successful multi-sided platform firms suggests that they got started in small markets first, and used trial and error to identify the correct technology and operations infrastructure in which to make further investments.[16] They avoided the trap of relying solely on the network economics literature or buzzword assumptions, and instead established successful buyer-seller transactions first, then making large scaling investments only when their platform had been tested.[17] EBay, one of the most successful Internet start-ups, started selling collectibles like the famous PEZ dispenser, and then moved into coins and stamps in 1995. It demonstrated a direct network effect, as the market making in collectibles became more attractive as its scale increased. Only after it was successful in collectibles did it expand into "early practicals" like computers, consumer electronics, and then books, movies, music and sports equipment (in 1997). Its third wave of "practicals," including motors, clothing, and home and garden didn't come until later (1999).[18]

We cited WebVan earlier as one of the notable train wrecks. Though gross margins in the retail grocery business are typically 5% or less, WebVan adopted a "get big fast" strategy and expanded into Chicago, Dallas, Washington D.C., and Seattle in November 1999. Despite rising expenses and shrinking margins, it moved into Atlanta in May 2000, one month before acquiring HomeGrocer.com for $1.2 billion. The company built enormous centralized automated warehouses and a hub and spoke distribution system in every market that it entered, on the assumption that the scale gave them cost advantages. Large trucks brought orders from the warehouses to staging areas, where they were transferred to small vans for delivery. This was a high fixed cost operation, and was very susceptible to underutilization. It did all this before really testing the business model. There was no way orders could come in fast enough to load the expensive infrastructure, so scale economies worked to the firm's detriment.

In contrast, Amazon.com has ventured slowly into online grocery delivery, conducting a small test in Mercer Island, Washington. Vauhini Vara writing in *The Wall Street Journal* reported on this experiment:[19]

> In the late '90s, a handful of online grocery-delivery companies such as Webvan Group and Homegrocer.com Inc. launched with fanfare but failed spectacularly when their sales didn't keep pace with the massive investment they had to make in infrastructure, like warehouses and delivery vans. Amazon, which is based in Seattle, was an early investor in

> Homegrocer.com. For its own service, dubbed Amazon Fresh, the company plans to store the groceries at a local refrigerated warehouse and to deliver them itself in "temperature-controlled totes" using 12 trucks of its own, but Mr. Berman declined to comment further on its investment in the service or on how many employees are working on it. "It's important that we're starting this very small, and that this is a test, and that we are going to evaluate our service closely and determine the best way to provide service," he said.

Scaling the business model

Network effects resemble economies of scale, and this is why many people confuse them, or at least are not careful about understanding the specific circumstances under which each applies. Economies of scale operate on the cost side and are inherently limited in their ability to create value (cost cannot go below zero, after all). Network effects operate on the willingness to pay side and don't necessarily have an upper bound. In both a business gets more profitable as it grows. So if it is not profitable at the onset, arguing that one of the effects exists is tantamount to saying, "We'll make it up on volume."

Managers build business models based on assumptions about scale economies, network effects, and the outcomes of system competitions. They would like to reach a tipping point and market dominance. Often the justification, especially with regard to Internet and Web 2.0 application service provider models (ASPs), is that competition is a "land grab" and that it is important to be first in the mind of potential consumers. But is it more important to get big fast, or is it more important to survive long enough to figure out the business model? Evans makes a telling observation:[20]

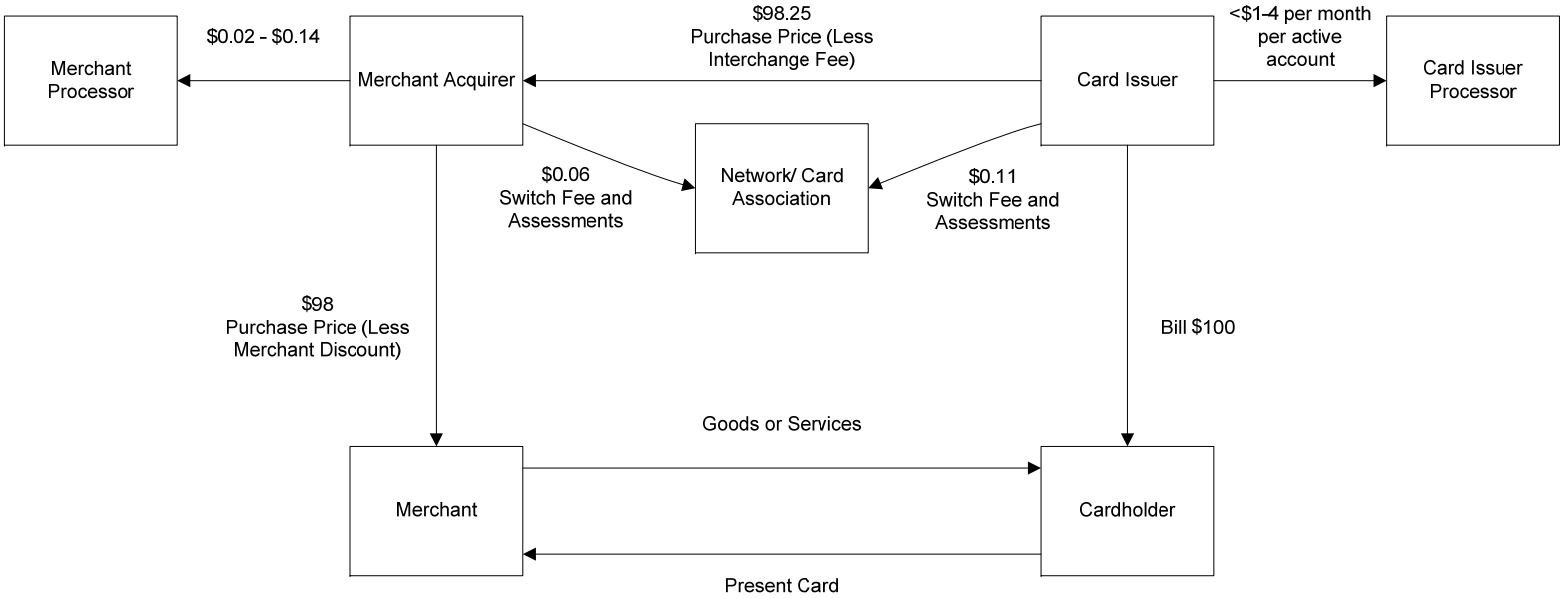
> Contrary to the traditional economic theory of network effects and the business advice based on that theory, there is no evidence that building up market share quickly is a recipe for market domination in platform industries most, if not all of which, are precisely those industries that economists have cited as having strong network effects. Many of the early entrants ultimately did not retain leadership positions: Diners Club in cards, Apple in personal computers, Apple in hand-held devices, and OnSale in online exchanges.

Exhibit 1 Examples of Multisided Platforms

Platform	Type of Platform	Side One	Side Two	Side that "Gets charged the Least"	Revenue model
Newspapers and magazines	Audience makers	Readers	Advertisers	Readers	Majority of revenue comes from selling advertising
Network television	Audience maker	Viewers	Advertisers	Viewers	Revenue comes primarily from advertisers
Internet Portals	Audience maker	Web surfer	Advertisers	Web surfers	The majority of revenues come from sale of advertising, though some portals charge subscription fees.
Operating System like Microsoft Windows	Demand coordinator	Application user	Application developer	Application developer	Most revenues generated from sale of packaged software
Video game consoles	Demand coordinator	Game player	Game developer	Both sides are significant sources of platform revenue	Game sales to end users and licensing fees from game developers are both significant revenue streams. Game console manufacturers typically sell consoles near or below marginal cost
Real Estate Broker	Market maker	Buyer	Seller	Buyer	Real estate brokers' income comes primarily from sales commissions
Apple iPhone	Demand coordinator	End Users	Application Developer	Application Developer	Hardware sales plus percentage of revenue for products sold through Apple AppStore
eBay	Demand coordinator Market maker	Buyer	Seller		Percentage of revenue fee for successful transactions

Source: Adapted from David S. Evans, "Some Empirical Aspects of Multi-sided Platform Industries," Review of Network Economics, Vol. 2, pp 191-209 (September 2003).

Exhibit 2 Pricing in the Payment Card System Value Network



- Merchant receives $98 ($100 purchase price less merchant discount)
- Merchant acquirer receives $2 merchant discount
- Merchant acquirer pays $1.75 Interchange Fee, $0.02 - $0.14 per transaction merchant processing fee, $0.03 per transaction network switch/operation fee and a ~1-2 bps assessment
- Card Issuer receives $1.75 interchange fee
- Card issuer pays $0.06 per transaction network switch/operation fee and a ~5-7 bps assessment
- Customer is billed $100

Source: Adapted by case writer from Tien-Tsing Huang, "The Electronic Payment Ecosystem," 2nd Edition, JPMorgan Computer Services & IT Consulting Equity Research (2007).

[1] Strictly speaking, what was cloned was the Basic Input Output System (BIOS) of the PC, which IBM had intended to keep proprietary. Once the BIOS was cloned, PCs could be manufactured that ran IBM software interchangeably.

[2] IBM eventually sold its disk drive operations to Hitachi, who renamed it Hitachi Global Storage Technologies HGST). HGST was acquired by Seagate Technologies in 2011.

[3] Elpida combined the memory operations of Hitachi and Mitsubishi.

[4] There were many motherboard manufacturers, initially in the United States but starting in the 1990s, this slice was dominated by Taiwanese makers like ASUSTeK, Microstar International, Gigabyte, Foxconn, and others.

[5] Compaq was sold to HP.

[6] IBM's PC operations were sold to Lenovo.

[7] Pet owners had no compelling reason to buy over the Internet, yet the wait for delivery put the company at a disadvantage. Ultimately the company offered free shipping, but its gross margin never moved into positive territory in any quarter of its existence.

[8] The Internet toy retailer had a peak market valuation of $8 billion before filing for bankruptcy protection in 2001, two years after its IPO.

[9] CNET described Boo.com as the "typical scenario … overhype, overfund, and over expand." The company burned through $135 million of venture capital money in 18 months, and had a staff of 400 people. It generated approximately $500,000 in sales before being liquidated.

[10] An excellent reference which provided many of the concepts discussed here is by Michael Katz and Carl Shapiro, "System Competition and Network Effects, " *Journal of Economic Perspectives*, Vol. 8, Number 2 (Spring 1994).

[11] Modular interfaces, especially when they become established as standards, facilitate this coordination. They enable customers to configure and assemble their own solutions for the job at hand. The new breed of netbook computers does this, using a web browser as the modular interface. The consumer coordinates with her desired applications by browsing the Internet.

[12] When companies like RCA Victor first introduced the phonograph, it had to create "software" in the form of pre-recorded music, so it built recording studios and entered the record business. Similarly when IBM first began selling large scale computer systems in the 1950s, it had to develop all of the software by itself. It was only as these markets matured and consumers could benefit from increased variety relative to integrated solutions did the modular boundary between hardware and software emerge. The computer hardware-software boundary was also transformed by the 1969 Department of Justice consent decree with IBM that forced the "unbundling" of hardware and software.

[13] Prof. Andrei Hagiu has done extensive research in this area, and has written an excellent background note, "Note on Multi-Sided Platforms: Economic Foundations and Strategy," HBS No. 9-709-484.

[14] David Becker, "Adobe's PDF-everywhere strategy," CNET News, September 30, 2003, http://news.cnet.com/2008-1012-5083805.html.

[15] "Shantanu Narayen on Adobe's Future Direction: Product Strategy for the Next Generation of the Web," *Knowledge@Wharton*, http://knowledge.wharton.upenn.edu/article.cfm?articleid=1741.

[16] Professors Thomas Eisenmann and Andrei Hagiu have written a note that describes effective strategies for building successful multi-sided platforms, "Staging Two-Sided Platforms," HBS No. 9-808-004. They describe

how firms might start out as a vendor on one side and then turn to the other side, or start as a "merchant buying from one side and reselling to the other.

[17] David Evans, "Some Empirical Aspects of Multi-sided Platform Industries, *Review of Network Economics*, Vol. 2, Issue 3, September, 2003.

[18] eBay Inc., 2001 Annual Report.

[19] Vauhini Vara, "Amazon Is Taking Small Steps Toward Grocery-Delivery Service," *The Wall Street Journal. (Eastern edition),* Aug 3, 2007, pg. B.4.

[20] David Evans, ibid.