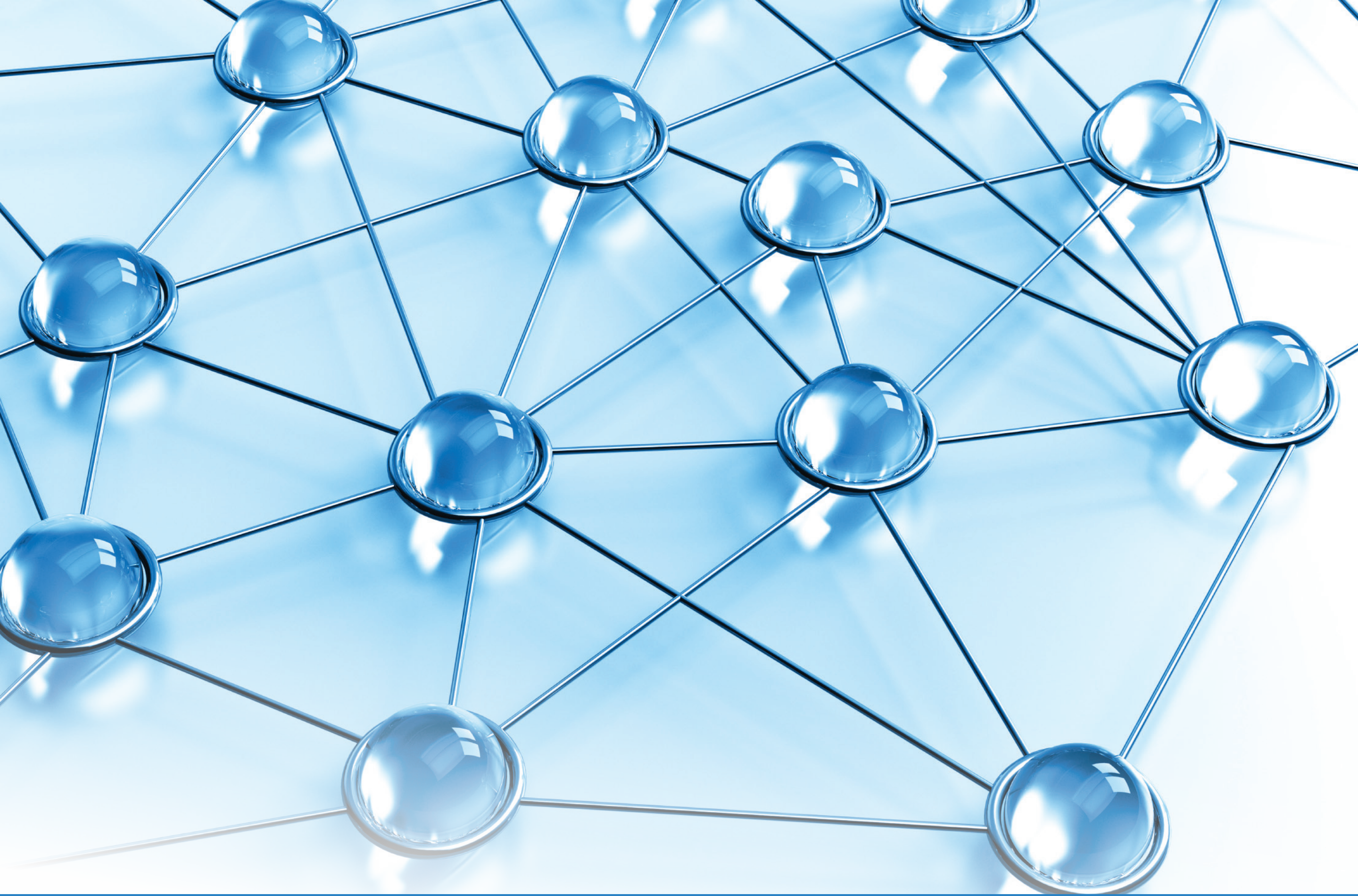


PLATFORM ECONOMICS:
Essays on Multi-Sided Businesses

David S. Evans

PLATFORM ECONOMICS:

Essays on Multi-Sided Businesses

David S. Evans

Competition Policy International (CPI) provides comprehensive resources and continuing education for the global antitrust and competition policy community. Created and managed by leaders in the competition policy community, CPI delivers timely commentary and analysis on antitrust and global competition policy matters through a variety of media and applications.

www.competitionpolicyinternational.com

Preface

Given the subject of this collection, there is some irony in how I've chosen to bring these essays to you. Publishing has traditionally been a two-sided model. Publishers get authors and readers together. They typically make their money by charging the reader and giving some fraction of the earnings to the author as royalties.

This 20th century model of publishing doesn't serve authors of academic books well. Often, publishers set the price of academic books relatively high, expecting to earn the greatest profits from libraries and a handful of aficionados. For most books that aren't aimed towards a popular audience, including most academic books, royalties are quite small. Optimistically, I might have been able to buy a pretty good new bicycle if I had published these essays in the traditional fashion, but I'd rather have more people read my work than collect the chump change from royalties.

Therefore, the two-sided publishing model fails in two ways: the author doesn't make much money, and the author doesn't get read by very many people. Moreover, most publishers in my experience are still using 20th century technology to produce and distribute books. It can take many months—if not years—from a book's conception to its appearance in a reader's hands.

And therein lies the paradox. In order to bring my work into the 21st century, I have decided to publish my collection of essays about two-sided markets in a one-sided way. I ditched the intermediary and chose to connect directly with likely readers. I'm sure some of you would prefer the feel of paper and leather but hopefully the price is right. It was easy for me to decide to make this volume free (a bit more on Amazon) because it cost almost nothing to produce and distribute it.

An earlier version of this book appeared in 2010. It consisted of a series of urls (website addresses) that took readers to the original papers which they could then download. I promised a real e-book in the early part of 2011. At least I got the year right which for an economist is pretty good.

David S. Evans

Acknowledgements

I would like to thank Richard Schmalensee for his collaboration on several of the articles and books on multi-sided platforms we have done over the last decade as well as Jean-Charles Rochet and Jean Tirole who have been generous with their comments and time. I also gratefully acknowledge research support from Microsoft and Visa for some of the articles. None of them necessarily agrees with anything I say of course. I also want to extend my appreciation to Almira Joy Bautisa and Dean Whitney of Aericon who led the effort to design and assemble this book.

Contents

Preface iii

Acknowledgements iv

Introduction vi

Part One: Economics of Multi-Sided Platform Businesses 1

Chapter 1: Industrial Organization of Two-Sided Platforms 2

Chapter 2: Some Empirical Aspects of Multi-Sided Platform Industries 30

Chapter 3: How Catalysts Ignite: The Economics of Platform-Based Startups 48

Part Two: Two-Sided Antitrust Economics 72

Chapter 4: Antitrust Economics of Multi-Sided Platforms 73

Chapter 5: Two-Sided Market Definition 135

Chapter 6: Defining Markets that Involve Two-Sided Platform 161

Part Three: Web Based Platforms 192

Chapter 7: Economics of the Online Advertising Industry 193

Chapter 8: The Online Advertising Industry: Economics, Evolution, and Privacy 226

Chapter 9: Web Economy, Two-Sided Markets, and Competition Policy 254

Chapter 10: Antitrust Issues Raised by Global Internet Economy 261

Part Four: Payment Cards 281

Chapter 11: More Than Money 282

Chapter 12: Interchange Fees and Their Regulation 314

Chapter 13: The Effect of Regulatory Intervention in Two-Sided Markets 347

Part Five: Software Platforms 375

Chapter 14: Invisible Engines 376

Chapter 15: Invisible Engines: Both Sides Now 382

Chapter 16: What's Next in Payments: Invisible Engines 410

Where Chapters Were Originally Published 428

Bibliography 430

Introduction

This volume collects a series of essays that I have written, sometimes with colleagues, over the last decade on businesses that create value by providing products that enable two or more different types of customers to get together, find each other, and exchange value. These businesses were called "two-sided markets" in the seminal paper by Jean-Charles Rochet and Jean Tirole that was first circulated in 2001.[1] I typically avoid this term since it tends to obscure the fact that we are talking about businesses rather than markets. I prefer the term "multi-sided platforms" because these businesses provide a place for customers to meet and interact and often support more than two interdependent types of customers. In writing for business audiences I use the term "catalyst" to denote the fact that these businesses create value that couldn't be had without bringing these customer types together.[2] The term "two-sided markets" has stuck, though, and I will use it here.

The Rochet and Tirole paper ignited work on two-sided markets in economics, law, and business. Several other papers, in circulation around 2000, touched on some of the interesting aspects of intermediaries[3] or on the increasingly widespread phenomenon of giving one product away for free to attract revenue from another product.[4] The Toulousians' contribution was fundamental because it recognized for the first time that a very diverse set of businesses were two-sided, presented an elegant economic model of them, and derived several robust aspects of these businesses including the critical importance of the price structure (the relative prices charged to the various types of customers) in their making money.

Economic theorists and empiricists were quickly attracted to this topic. The mill of articles and dissertations has flourished ever since.[5] It soon became apparent that this new area had important implications for antitrust. Many competition authorities took notice and law review

---

[1] Jean-Charles Rochet and Jean Tirole, "Platform Competition in Two-Sided Markets," Journal of the European Economic Association 1, no. 4 (2003): 990-1209.

[2] David S. Evans and Richard Schmalensee, Catalyst Code: The Strategies behind the World's Most Dynamic Companies (Cambridge, MA: Harvard Business School Press, 2007).

[3] Bernard Caillaud and Bruno Jullien, " Chicken & Egg: Competition among Intermediation Service Providers," RAND Journal of Economics, The RAND Corporation, vol. 34(2), pages 309-28, Summer.

[4] Geoffrey Parker and Marshall Van Alstyne (2005). "Two-Sided Network Effects: A Theory of Information Product Design." Management Science, Vol. 51, No. 10.

[5] With the recent important work of Glen Weyl we have moved on to Two-Sided Markets 2.0. See Glen Weyl, "A Price Theory of Multi-Sided Platforms," American Economic Review, 2010, 100(4).

articles addressed various aspects of these newly-recognized business forms.[6] The corporate world seized on this new field. Strategy articles and courses focused on these platforms began appearing.[7] Companies found the subject eye opening, including firms that were multi-sided platforms but had not quite understood the ramifications.[8]

My contributions to this literature are presented in this volume. Part I presents background pieces on the economics of multi-sided platforms and industries in which these platforms are common. Part II examines the antitrust economics of two-sided markets including the difficult problem of defining the boundaries of competition. Part III comprises several papers that apply two-sided market analysis to web-based businesses. Part IV does the same for payment cards which is the industry that attracted much of the early two-sided analysis— in part because this framework was helpful for understanding the hotly debated issue of interchange fees. Part V collects several article and book chapters on software platforms. These platforms have become especially important in the last several years because they are now the basis for revolutionary developments with mobile devices (e.g. the iPhone and Android), social networking (Facebook in particular), and payments (PayPalX). The essays are published as originally written (usually, in fact, whatever version I could make freely available).

When the theory of two-sided markets was first introduced it was common to hear at least two complaints. The first was that there was nothing new—from economists who suggested that it was just the indirect network effects wine in new bottles or antitrust analysts who commented that it had all been considered before in advertising cases. The second was that it was a theory of everything, and therefore nothing, since everything seems to be two-sided.

There's some truth, of course to both arguments. Indirect network effects are usually essential to understanding two-sided markets. But the network effects literature missed the importance of these effects for a very diverse group of industries among other things; the literature spent a lot of time on fax machines and video standards but not so much on more general business issues such as pricing or industries such as shopping malls that did not obviously have indirect network effects. Two-sided market analysis is at least a richer and more subtle wine.

One of the problems with two-sided market analysis is that it is hard to find formal limiting principles. But that isn't uncommon in economics. Sometimes a two-sided market perspective is highly informative while other times it isn't. It matters when it matters. What's now very clear, with the benefit of a decade of work, is that the study of multi-sided platforms has provided valuable insights to economists, policymakers and business people, as I hope the chapters in this volume demonstrate.

---

[6] OECD Organization for Economic Co-operation and Development ("OECD"), Two-Sided Markets, DAF/COMP (2009) 20, 28.

[7] See, for example, Thomas Eisenmann "Managing Networked Businesses: Course Overview." Harvard Business Online , 2006;Thomas Eisenmann, Geoffrey Parker, and Marshall Van Alstyne, "Strategies for Two-Sided Markets." Harvard Business Review, 2006.

[8] Companies in the payment card industry, for example, have fundamentally changed how they think about that business as a result of the two-sided market concepts.

Part One

ECONOMICS OF MULTI-SIDED PLATFORM BUSINESSES

CHAPTER 1

The Industrial Organization of Markets with Two-Sided Platforms

David S. Evans and Richard Schmalensee

ABSTRACT

Many diverse industries are populated by businesses that operate "two-sided platforms." These businesses serve distinct groups of customers who need each other in some way, and the core business of the two-sided platform is to provide a common (real or virtual) meeting place and to facilitate interactions between members of the two distinct customer groups. Platforms play an important role throughout the economy by minimizing transactions costs between entities that can benefit from getting together. In these businesses, pricing and other strategies are strongly affected by the indirect network effects between the two sides of the platform. As a matter of theory, for example, profit-maximizing prices may entail below-cost pricing to one set of customers over the long run and, as a matter of fact, many two-sided platforms charge one side prices that are below marginal cost and are in some cases negative. These and other aspects of two-sided platforms affect almost all aspects of antitrust analysis—from market definition, to the analysis of cartels, single-firm conduct, and efficiencies. This chapter provides a brief introduction to the economics of two-sided platforms and the implications for antitrust analysis.

I. INTRODUCTION

Many diverse industries are populated by businesses that operate "two-sided platforms." These businesses serve distinct groups of customers who need each other in some way, and the core business of the two-sided platform is to provide a common (real or virtual) meeting place and to facilitate interactions between members of the two distinct customer groups. Two-sided platforms are common in old-economy industries such as those based on advertising-supported media and new-economy industries such as those based on software platforms and web portals. They play an important role throughout the economy by minimizing transactions costs between entities that can benefit from getting together.

In these businesses, pricing and other strategies are strongly affected by the indirect network effects between the two sides of the platform. As a matter of theory, for example, profit-maximizing prices may entail below-cost pricing to one set of customers over the long run and, as a matter of fact, many two-sided platforms charge one side prices that are below marginal cost and are in some cases negative. These and other aspects of two-sided platforms affect almost all aspects of antitrust analysis—from market definition, to the analysis of cartels, single-firm conduct, and efficiencies.[1]

This paper provides a brief introduction to the economics of two-sided platforms and the implications for antitrust analysis.

Two-sided platforms were first identified clearly in pioneering work by Jean Charles Rochet and Jean Tirole, which began circulating in 2001.[2] A significant theoretical and empirical literature quickly emerged, and the subject has become a very active area of research in economics.[3] For the purposes of this paper, it is helpful to clarify some terminology that is used in the economics literature and which sometimes causes confusion. Rochet and Tirole used the term "two-sided markets" to refer to situations in which businesses cater to two interdependent groups of customers. The term "market" was meant loosely and does not refer to how that term is often used in antitrust. This paper refers to "two-sided platforms" but it is synonymous with "two-sided markets" as used in much of the economics literature. How to determine what market a two-sided platform competes in, from an antitrust perspective, is one of the questions considered here.[4] Two-sided platforms often compete with ordinary (single-sided) firms and sometimes compete on one side with two-sided platforms that serve a different second side.

II. ECONOMIC BACKGROUND ON TWO-SIDED PLATFORMS

A heterosexual, singles-oriented club offers some intuition on the economics of two-sided platforms. A nightclub, such as Bungalow 8 in Manhattan, provides a platform where men and women can meet and search for interactions and potentially dates. The club needs to get two groups of customers on board its platform to have a service to offer either one: it needs to get both men and women to come. Moreover, the relative proportion of men and women matters. A singles club with few women will not attract men, and a club with few men will not attract women. Pricing is

---

1 *See* David S. Evans, *The Antitrust Economics of Multi-Sided Platform Markets*, 20 YALE J. ON REG. 325 (2003) and Julian Wright, *One-Sided Logic in Two-Sided Markets*, 3 REV. OF NETWORK ECON. 44 (2004).

2 Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. EUR. ECON. ASS'N 990 (2003). Some of the key issues were identified in the context of payment cards in an important contribution Wlliam F. Baxter, *Bank Exchange of Transactional Paper: Legal and Economic Perspectives*, 26 J.L. & ECON. 541 (1983). There are also literatures for particular industries that also provide precursors.

3 *See* Conference on Competition Policy in Two-Sided Markets (Institute d'Economie Industrielle, U. Toulouse) (Jun. 29 - Jul. 1, 2006), *available at* http://idei.fr/doc/conf/tsm/programme.pdf.

4 Although, for the most part, we will use the term two-sided platform the reader should note that some platforms have more than two distinct groups of customers. Digital media platforms, for example, often have four: users, developers, hardware makers, and content providers.

one way to get the balance right. The club might want to offer women a break if they are in short supply (through a lower price or free drinks). Or it might want to ration the spots to ensure the appropriate number of women; popular clubs typically have queues waiting outside, and women are picked out of line disproportionately.

The dating club example motivates the informal definition of a two-sided platform that we introduced in the beginning paragraph. There are two groups of customers—men and women. Members of each group value members interacting with members of the other group. And the platform provides a place for them to get together and interact. By doing so it enables members of these two groups to capture various benefits from having access to each other.

In their 2006 paper, Rochet and Tirole have proposed a formal definition:

> "A market[5] is two-sided if the platform can affect the volume of transactions by charging more to one side of the market and reducing the price paid by the other side by an equal amount; in other words, the price structure matters, and platforms must design it so as to bring both sides on board."[6]

To satisfy this definition, "the relationship between end-users must be fraught with residual externalities" that customers cannot sort out for themselves.[7] That is clear in the case of the dating environment. In contrast, in the textbook wheat market there are no externalities connecting buyers and sellers, and the price structure doesn't matter: a tax on wheat levied on buyers has the same effect on quantity as the same tax levied on sellers.

In addition, it must not be possible for the two sides to arbitrage their way around the price structure chosen by the platform. Men and women, for example, want to be able to search for dates among a large number of opposites. It is hard to conceive of a practical mechanism for women to reward men who come to a singles club but who they reject. Likewise, for the other two-sided platform industries we consider it is difficult, if not impossible, for customers on one side to make side payments to customers on the other side. As a result the platform owner can institute a pricing structure to harness indirect network effects, and it is not feasible for customers

Think of two-sided platforms as arising in situations in which there are externalities and in which transactions costs, broadly considered, prevent the two sides from solving this externality directly.

[5] Note that the word market below is being used in the loose manner that is the custom among economists and not in the antitrust sense. The Rochet-Tirole definition would be more precise if it said "A two-sided platform business exists if …."

[6] Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report,* RAND J. Econ. (Autumn 2006).

[7] As a result a necessary condition for a market to be two-sided is that the Coase theorem does not apply to the transaction between the two sides. For more details, see Rochet & Tirole (2006), *id.*

to defeat this pricing structure through arbitrage. Generally, one can think of two-sided platforms as arising in situations in which there are externalities and in which transactions costs, broadly considered, prevent the two sides from solving this externality directly. The platform can be thought of as providing a technology for solving the externality in a way that minimizes transactions costs.

It is helpful to review four different types of two-sided platforms: exchanges, advertiser-supported media, transaction devices, and software platforms.[8]

A. Exchanges

Exchanges have two groups of customers, who can generally be considered "buyers" and "sellers." The exchange helps buyers and sellers search for feasible contracts—that is where the buyer and seller could enter into a mutually advantageous trade—and for the best prices—that is where the buyer is paying as little as possible and the seller receiving as much as possible. (In organized exchanges, As a result a necessary condition for a market to be two-sided is that the Coase theorem does not apply to the transaction between the two sides. For more details, *see* Rochet & Tirole (2006), id. such as the New York Stock Exchange, it is often more useful to think of the two sides as liquidity providers—specialists or market-makers who quote prices to both buyers and sellers and thus bring liquidity to the market—and liquidity consumers—ordinary customers who accept liquidity providers' offers.[9]) We use the term buyers and sellers here loosely. The term, "exchanges," covers various match-making activities such as dating services and employment agencies. It also covers traditional exchanges such as auction houses, internet sites for business-to-business, person-to-business, and person-to-person transactions, various kinds of brokers (insurance and real estate) and financial exchanges for securities and futures contracts. Finally, exchanges include a variety of businesses that provide brokerage services. These include publishers (readers and authors), literary agents (authors and publishers), travel services (travelers and travel-related businesses), and ticket services (people who go to events, and people who sponsor events).

Exchanges provide participants with the ability to search over participants on the other side and the opportunity to consummate matches. Having large numbers of participants on both sides increases the probability that participants will find a match. Depending on the type of exchange,

---

[8] For discussion, see David S. Evans, Andrei Hagiu, & Richard Schmalensee, Invisible Engines: How Software Platforms Drive Innovation and Transform Industries, ch. 3 (MIT Press 2006). We refer there to software platforms more generally as shared input facilities. Armstrong uses the term "competitive bottle-necks" to refer to certain shared-input facilities. Although his discussion is analytically sound, his term is pejorative and has a meaning in competition law that differs from the one he assigns to it. See Mark Armstrong, Competition in Two-Sided Markets (EconWPA, working paper, 2005).

[9] Bernhard Friess & Sean Greenaway, *Competition in EU Trading and Post-Trading Service Markets*, 2 Competition Pol'y Int'l (2006).

however, a larger number of participants can lead to congestion. That is the case with physical platforms such as singles clubs or trading floors. Moreover, participants may derive some value from having the exchange prescreen participants to increase the likelihood and quality of matches.

Some exchanges charge only one side. For example, only sellers pay directly for the services provided by eBay. This is also true for real-estate sales in the United States. Other exchanges charge both sides, although the prices may bear little relation to side-specific marginal costs. Internet matchmaking services charge everyone the same, for instance, while, as we mentioned, physical dating environments sometimes charge men more than women. Auction houses charge commissions to buyers and sellers. Insurance brokers historically charged both insurance customers and insurance providers in some types of transactions (some have agreed not to charge both as a result of settlements of lawsuits brought by the New York State Attorney General).

B. Advertising-Supported Media

Advertising-supported media such as magazines, newspapers, free television, and web portals are based on a two-sided business model. The platform either creates content (newspapers) or buys content from others (free television). The content is used to attract viewers. The viewers are then used to attract advertisers. There is a clear indirect network effect between advertisers and viewers—advertisers value platforms that have more viewers; the extent to which viewers value advertisers is the subject of more debate but we suspect that viewers value advertisers more than they might admit.[10]

Most advertising-supported media earn much of their revenues—and probably all of their gross margin—from advertisers.[11] Print media are often provided to readers at something close

[10] *See, e.g.,* James M. Ferguson, *Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition*, 26 J.L. & Econ. 637 (1983) ("Readership studies show that advertising, especially retail advertising, is considered as important as, or more important than, editorial content.") and R.D. Blair & R.E. Romano, *Pricing Decisions of the Newspaper Monopolist*, 59 Southern Econ. J. 731 (1993) ("circulation demand rises with increases in the quantity of advertising").

Other studies have shown that, unlike Americans, readers in certain European countries are averse to advertising. *See, e.g.*, Nathalie Sonnac, *Readers' Attitudes Toward Press Advertising: Are They AdLovers or Ad-Averse?*, 13 J. Media Econ. 249 (2000). On the other hand, TiVo and other related products that permit ad avoidance and deletion are very popular currently, with one study citing that TiVo viewers skip about 60 percent of commercials. *See A Farewell to Ads?*, The Economist, Apr. 15, 2004.

[11] In a two-sided platform there is no rigorous way to define the profit "earned" by one side or the other. Not only are there typically costs that are common to both sides (the floor of the New York Stock Exchange, for instance), outlays that build business on one side of the market (via product enhancement, say) will also tend, via the externality, to build business on the other side. By "gross margin" we mean the difference between revenue and the variable costs, if any, that depend entirely on the volume on only one side of the market. The cleanest examples of such a cost would be the manufacturing costs of videogame consoles or the marginal printing costs of newspapers or yellow page directories.

to or below the marginal cost of printing and distribution.[12] In some cases—such as yellow page directories and some newspapers—they are provided for free. Free television is just that. And most web portals—Google and Yahoo for example—receive revenue only from advertisers.

C. Transaction Systems

Any method for payment works only if buyers and sellers are willing to use it. Humans switched from barter when they were agreed on a standard medium for exchange—such as metallic coins or seashells. Governments facilitated this by ensuring the integrity of coins (to various degrees) and by using government issued coinage for buying and selling. Cash, which has no intrinsic value in most modern economies, provides a payment platform because buyers and sellers expect that other buyers and sellers will use it. Of course the government facilitates this with various laws and through its own buying and selling activities.

For-profit transaction systems are based on the same principles although they have challenges that governments—which at least in principle can create a platform by fiat—do not necessarily have. Although bank checks and travelers' checks are also examples of for-profit transaction systems, we focus on payment cards, which have been the subject of significant competition policy scrutiny in many countries.

Diners Club started the first two-sided payment system in 1950. Before then stores issued payment cards to their customers for use only at their stores. Diners Club began by getting a set of restaurants to agree to take its card for payment; that is to agree to let Diners Club reimburse the restaurant for the meal tab and then in turn collect the money from the cardholder. It also persuaded individuals to take its card and use it for payment. Starting with a small base in Manhattan it grew quickly throughout the United States and other countries.

Diners Club initially charged restaurants seven percent of the meal tab; cardholders had to pay an annual fee, which was offset in part by the float they received as a result of having to pay their bills only once a month. As a result Diners Club earned most of its revenue—and most likely all of its gross margin—from merchants. Other entrants into the charge and debit card businesses have followed this same approach. Determining who pays in the case of credit cards is a bit more complicated since that product bundles a transaction feature (for which the cardholder pays little) and a borrowing feature (for which the cardholder incurs finance charges). However, it is safe to say that merchants are the main source of revenue for credit cards held by people who do not revolve balances.

American Express, Discover, and, until its recent absorption into MasterCard, Diners Club, set prices to merchants—the merchant discount, which gives rise to a positive variable transaction price—and to cardholders—annual fees and various rewards which may give rise to negative

[12] Blair & Romano, *supra* note 10.

variable transaction prices. Card associations such as MasterCard and Visa have been examples of cooperative two-sided platforms. For a transaction to be consummated there has to be an agreement on the division of profits and the allocation of various risks between the entity that services the merchant and the entity that services the cardholder. Most card associations set this centrally as, in effect, a standard contract between the businesses that service the two sides. Typically, they agree that the entity that services the merchant pays a percentage of the transaction—the "interchange fee"—to the entity that services the cardholder. This fee ultimately determines the relative prices for cardholders (issuers obtain a revenue stream which they compete for) and merchants (acquirers pass the cost of the interchange fee onto merchants). This centrally set fee has been the subject of litigation and regulatory scrutiny, as we discuss below.[13]

D. Software Platforms

A software platform provides services for applications developers; among other things, these services help developers obtain access to the hardware for the computing device in question. Users can run these applications only if they have the same software platform as that relied on by the developers; developers can sell their applications only to users that have the same software platform they have relied on in writing their applications.

Software platforms are central to several important industries. These include personal computers (e.g., Apple, Microsoft); personal digital assistants (e.g., Palm, Treo); 2.5G+ mobile telephones (e.g., Vodafone, DoCoMo); video games (e.g., Sony PlayStation, Xbox); and digital music devices (e.g., Creative Zen Micro, Rio Carbon). With the exception of video games, the software platform owners make most of their revenue, and all of their gross margin, from the user side; developers generally get access to platform services for free, and they obtain various software products that facilitate writing applications at relatively low prices. Videogame console manufacturers, on the other hand, typically receive most of their gross margin from licensing access to the software and hardware platforms to game developers; they sell the videogame console at close to or below manufacturing cost.

Software platforms facilitate a market for applications by reducing duplicative costs. Application programs need to accomplish many similar tasks. Rather than each application developer writing the code for accomplishing each task the software platform producer incorporates code into the platform. The functions of that code are made available to application developers through an application program interface (API). The user benefits from this consolidation as well since it reduces the overall amount of code required on the computer, reduces incompatibilities between

[13] David S. Evans & Richard Schmalensee, The Economics of Interchange Fees and Their Regulation: An Overview (MIT Sloan, Working Paper, 2005), *in Interchange Fees in Credit and Debit Card Industries* 73-120 (Federal Reserve Bank of Kansas City, 2005).

programs, and reduces learning costs.[14] An important consequence of this reduction in cost is an increase in the supply of applications for the platform, an increase in the value of the software platform to end users, and positive feedback effects to application developers.

E. Methods for Minimizing Transactions Costs

The fundamental role of a two-sided platform in the economy is to enable parties to realize gains from trade or other interactions by reducing the transactions costs of finding each other and interacting. Two-sided platforms do this by matchmaking, building audiences, and minimizing costs. Different platforms engage in these activities to different degrees. Software platforms are mainly about minimizing duplication costs, advertising-supported media in mainly about building audiences, and exchanges are mainly about matchmaking. But they all seem to engage in each to some degree. All platforms help reduce costs by providing a virtual or physical meeting place for customers. We will see that these platforms all minimize transactions costs by through matchmaking, audience making, and cost minimization through the elimination of duplication.[15]

MySpace provides an example of how a two-sided platform engages in all three functions. It is a popular internet site where young people can post their profiles and develop networks of friends. It provides matchmaking between the people who sign up as well as the advertisers who would like to meet them. It builds audiences for advertisers as well as members—particularly musicians—who want to make themselves known. And it reduces the costs to people of getting together by providing a common meeting place.

III. ECONOMIC PRINCIPLES

The theoretical economics literature on two-sided platforms is relatively new. Economists have derived many results based on stylized models that apply to some of the industries described above. The precise results are sensitive to assumptions about the economic relationships among the various industry participants. Even for these special cases it has turned out to be challenging to derive results without making further assumptions about the precise nature of the demand, cost, and indirect network effects relationships.[16] Nevertheless, several principles have emerged that seem to be robust. They appear to depend only on the assumptions that the platform has two groups of customers, that there are indirect network externalities, and that the customers cannot solve these externalities themselves.

---

[14] *See* Evans, Hagiu, & Schmalensee, *supra* note 8.

[15] *See* David S. Evans & Richard Schmalensee, Catalyst Code: The Strategies Behind the World's Most Successful Companies (Harvard Business School Press 2007).

[16] That is, the models are based on assuming particular functional forms—e.g. linear—for relationships.

A. Pricing

To see the intuition behind pricing consider a platform that serves two customer groups ***A*** and ***B***. It has already established prices to both groups and is considering changing them.[17] If it raises the price to members of group ***A*** fewer ***A***s will join. If nothing else changed the relationship between price and the number of ***A***s would depend on the price elasticity of demand for ***A***s. Since members of group ***B*** value the platform more if there are more ***A***s fewer ***B***s will join the platform at the current price for ***B***s. That drop-off depends on the indirect network externality which is measured by the value that ***B***s place on ***A***s. But with fewer ***B***s on the platform, ***A***s also value the platform less leading to a further drop in their demand. There is a feedback loop between the two sides. Once this effect is taken into account, the effect of an increase in price on one side is a decrease in demand on the first side because of the direct effect of the price elasticity of demand and on both sides as a result of the indirect effects from the externalities.

A few equations will make this point more sharply for readers familiar with the concept of elasticity. The situation described just above can be summarized by two demand functions: $Q^A = D^A(P^A, Q^B)$ and $Q^B = D^B(P^B, Q^A)$. The first of these gives participation by members of group ***A*** as a function of the price charged to group ***A*** and participation by group ***B***, and the second gives participation by members of ***B*** similarly. Let $e^I = -(\partial D^I/\partial P^I)(P^I/Q^I)$, for $I = A,B$. These are the own-price elasticities for each group, holding constant *participation* by the other (i.e., ignoring the externalities linking the two groups). Let $\theta^I_J = (\partial D^I/\partial Q^J)(Q^J/Q^I)$ for $I, J = A, B$ and $I \neq J$. These elasticities measure the strengths of the externalities connecting the two groups. In the normal two-sided case, both would be expected to be positive. Finally, let $E^I = -(dQ^I/dP^I)(P^I/Q^I)$ for $I = A, B$. These are the ordinary own price elasticities, computed assuming other prices remain constant but allowing participations (quantities) to vary. Differentiating both demand functions totally with respect to either price, and solving, yields:

$$E^I = e^I/(1 - \theta^I_J\theta^J_I);\ I, J = A,B;\ I \neq J.$$

Even if the ***A***s are not particularly price-sensitive, and as long as the externalities between the groups are strong (in either direction!), participation by group ***A*** may be highly sensitive to the price its members are charged, and similarly for group ***B***. Even a small response by group ***A*** to a price change will trigger a response by group ***B***, which in turn will produce a response by ***A***, and so on. (The equation above assumes that these response sequences converge.)

The platform of course would like to find the prices that maximize its profits by taking these same sorts of considerations into account. For a single-sided business that would occur by selecting

[17] To keep matters simple we consider the case where each side is charged a membership fee as in Mark Armstrong, Competition In Two-Sided Markets (EconWPA, Working Paper, 2005). More generally, platforms are natural businesses for two-part tariffs involving an access fee and a usage fee.

the output at which marginal revenue equals marginal cost and then charging the corresponding price for this quantity from the demand curve. (This equilibrium is often described by the Lerner formula that says that the price marginal-cost margin equals the inverse of the own-price elasticity of demand.) For two-sided platforms three results appear to be robust:

1. The optimal prices depend in a complex way on the price sensitivity of demand on both sides, the nature and intensity of the indirect network effects between the two sides, and the marginal costs that result from changing output of each side.
2. The profit-maximizing, non-predatory price for either side may be below the marginal cost of supply for that side or even negative.
3. The relationship between price and cost is complex, and the simple formulas that have been derived for single-sided markets do not apply.

For many platforms it is possible to charge two different kinds of prices: an access fee for joining the platform and a usage fee for using the platform. Although these are interdependent, one can think of the access fee as mainly affecting how many customers join the platform and the usage fee as mainly affecting the volume of interactions between members of the platform. Most software platforms charge access fees to users—they have to license the software platform but then can use it as much as they want—and do not charge access or usage fees to developers. Videogame console vendors, though, charge a usage fee to game developers—a royalty based on the numbers of games that are sold; users pay this usage fee indirectly through their purchase of games for the console. Payment card systems generally charge merchants a usage fee but no access fee. Cardholders may pay an access fee (the annual card fee); they often pay either no usage fee or a negative one (to the extent they receive rewards based on transactions volume).

The profit-maximizing reliance on access versus usage fees depends on many factors including the difficulty of monitoring usage and the nature of the externality between the two sides. Cardholders care about card acceptance, for instance, while merchants care about usage. It thus seems sensible not to charge merchants for access and not to charge consumers for usage.

The empirical evidence suggests that prices that are at or below marginal cost are common for two-sided platforms. Table 1 summarizes some relevant evidence.

B. Design Decisions

Two-sided platforms are in the business of encouraging customers to join their platforms and stimulating them to interact with each other once they have joined. They design their platforms with this in mind. This can lead to decisions that in a narrow sense harm one side.

A simple example is a shopping mall. Shoppers would prefer to get to stores in the least amount of time. Merchants would like to maximize the amount of foot traffic outside their stores and there-

TABLE 1[18] Examples of two-sided pricing structures

Industry	Side	Access	Usage
Heterosexual Dating Clubs	Men	√	√
	Women	√	√
DoCoMo i-Mode	User	√	√
	Content-Provider	Ø	√
U.S. Real Estate Brokers	Seller	Ø	√
	Buyer	Ø	Ø
Magazines	Reader	√ (≤MC)	Ø
	Advertiser	Ø	√
Shopping Malls	Shopper	–	Ø
	Store	√	Ø
PC Operating Systems	User	√	Ø
	Developer	√ (<MC)	Ø
Video Game Consoles	Player	√ (≤MC)	Ø
	Game Developer	√ (<MC)	√
Payment Card Systems	Merchant	Ø	√
	Cardholder	√ (<MC)	Ø

Note: √ and Ø indicate that the entity either pays or does not pay, respectively, for either access or usage of the two-sided platform. Items parentheses indicate where marginal does not below respectively, cost pricing prevalent usage of two-sided of a two-sided platform.

[18] This table shows pricing structures that are common in these industries. In many cases, fees will differ from these pricing structures. For example, some clubs offer free entry to women, some magazines offer free subscriptions, some videogame players pay fees for on-line play, and some payment cardholders do not pay fees for their cards and/or get usage based rewards. For dating clubs, usage fees for men and women refer to fees for drinks in the club. For real estate, the usage fee for sellers refers to the fee for selling a house; there is typically no fee for using the system to list or show a house. For shopping malls, the negative usage fee for shoppers refers to the free parking that is commonly available. For videogame consoles, players do not pay a fee for using the console, although they do pay for video games to the game developer (which in some cases is the same firm that makes the console and in other cases pays a royalty to the console manufacturer). For payment cards, cardholders are also subject to penalty fees, such as for exceeding credit limits or for late payments; we have not included these fees in the table.

fore the number of potential shoppers. Shopping malls are sometimes designed to encourage shoppers to pass by many stores (e.g., by putting the up and down escalators at different ends of the mall).

Advertising-supported media are another obvious example. Viewers would like to gain access to the content—and perhaps even the advertisements of their choice—in the most convenient way. Some magazines are laid out to make it difficult even to find the table of contents or to find the continuation of an article without thumbing through many advertisements. Television watchers might benefit from having advertisements clustered at the beginning or the end of each program, but television providers (in the United States, at least) typically intersperse the advertisements and precede them perhaps with a cliffhanger to discourage viewers from taking a long break.

Two-sided platforms may also bundle features that directly benefit side A but harm side B (putting aside the indirect externalities from increasing the participation of side A).[19] All software platforms include features for example that do not benefit most users. However, some developers value each of these features and in particular value knowing that any user of the software will have that feature and therefore be able to run its applications. All payment card systems require merchants that take their cards for payment to take any of their cards for payment, regardless of who presents it or which entity issued it. Some merchants would benefit from being selective—taking cards only from people who lack cash, for example. But this would reduce the confidence that cardholders have that their cards will be taken at stores that display the acceptance mark. (We will see later that special cases of these requirements, linking acceptances of credit and debit cards, have given rise to tying claims. This paragraph is not meant to suggest that tying could not be used in an anticompetitive way by two-sided platforms but rather to point out that there is an additional efficiency explanation for at least one aspect of this practice that does not arise in one-sided businesses.)

C. Rules and Regulations

Given that platforms promote interactions between customers and seek to harness indirect network externalities it should come as no surprise that two-sided platforms have an incentive to devise rules and regulations that promote these externalities and limit negative externalities between customers. The most sophisticated rules and regulations may be those employed by exchanges. All exchanges have rules against "front-running," for instance. This practice occurs when a broker receives a large purchase order from a customer, first buys on his own account, and then executes the customer order, which drives the price up slightly, and then sells on his own account and pockets the resulting profit. Banning this practice directly harms brokers, but it makes buyers and sellers more confident that they are getting the best price possible, and thereby boosts volume on the exchange.

Cooperative two-sided platforms have further need for rules and regulations because the behavior of their members can affect the value of the two-sided platform as a whole. Visa, for ex-

[19] *See* Rochet & Tirole (2006), *supra* note 6.

ample, has rules that govern the appearance of cards issued by members, to provide some uniformity for the common brand, as well as to prevent members from using the brand inappropriately. The system also has rules that address disputed transactions. Acquirers would have an incentive to favor their customers (merchants) in a dispute while issuers would favor their customers (cardholders). The system's rules attempt to find a balance between these competing interests, to increase the attractiveness of the system as a whole.

Cooperative two-sided platforms have further need for rules and regulations because the behavior of their members can affect the value of the two-sided platform as a whole.

IV. INDUSTRIAL ORGANIZATION OF MARKETS WITH TWO-SIDED PLATFORMS

Casual empiricism shows that industries with two-sided platforms are quite diverse. We explain some of the basic determinants of this heterogeneity from a theoretical perspective and then document aspects of it by surveying industries in which two-sided platforms are central.

A. Determinants of Platform Size and Structure

Five fundamental factors determine the relative size of competing two-sided platforms. Table 2 summarizes the factors we discuss below and their effect on size (with a "+" indicating that there is a positive association between size and the factor).

1. Indirect Network Effects

Indirect network effects between the two sides promote larger and fewer competing two-sided platforms. Platforms with more customers of each group are more valuable to the other group. For example, more users make software platforms more valuable to developers and more developers make software platforms more valuable to users. These positive-feedback effects make platforms

TABLE 2 Determinants of industry structure

Cause	Effect on Size/Concentration
Indirect network effects	+
Scale economies	+
Congestion	–
Platform differentiation	–
Multi-homing	–

with more customers on both sides more valuable to both sets of customers. To take another example, a payment card system whose cards are taken at more merchants is more valuable to card users—that is why we see card systems touting their acceptance ("MasterCard: No card is more accepted.") in consumer advertisements.

If there were no countervailing factors, we would expect that indirect network effects would lead two-sided platforms to compete ***for*** the market. First movers would have an advantage, all else being equal. We would have the familiar story that the firm that obtains a lead tends to widen that lead as a result of positive feedback effects and therefore wins the race for the market.[20] Other firms could compete with this advantage only if they offered consumers on either side something that offset the first mover's size advantage.

Indirect network effects may decline with the size of the platform. For example, the probability of finding a match increases at a diminishing rate with the number of individuals on either side (buyers or sellers, men or women).[21] At some point positive externalities from more participants may turn into negative externalities in the form of congestion as discussed below.

2. Economies and Diseconomies of Scale

For many two-sided platforms there would appear to be significant fixed costs of providing the platform. This should lead to scale economies over some range of output. For example, card payment systems have to maintain networks for authorizing and settling transactions for cardholders and merchants (and for their proxies—issuers and acquirers—in the case of association-based payment systems such as MasterCard). The costs of developing, establishing, and maintaining these networks are somewhat independent of volume. To take another example, there is a fixed cost of developing a software platform but a low marginal cost of providing that platform to developers and end users. In some cases the scale economies may mainly operate on one side. For example, there are scale economies in providing newspapers to readers (there is a high fixed cost of creating the newspaper and a relatively low marginal cost of reproducing and distributing it) but not in providing space to advertisers. Lastly, some physical platforms such as trading floors and singles clubs have scale economies at least in the short run, up to their capacity levels.

Diseconomies may set in at some point for various reasons on one or both sides. For example, to persuade existing end users to replace (i.e., upgrade) their existing software platforms software, platform vendors have to add features and functionality. Many of these improvements may be designed to encourage application developers to write new or improved applications for the platform that in turn benefit end users. However, as software platforms have gotten larger and more

---

[20] *See, e.g.,* David S. Evans & Richard Schmalensee, *A Guide to the Antitrust Economics of Networks*, 10 Antitrust Mag. 36 (1996) and Carl Shapiro & Hal R. Varian, Information Rules: A Strategic Guide to the Network Economy (Harvard Business School Press 1999).

[21] *See* Evans, *supra* note 1.

complex, it has become more expensive and time consuming to add features and functionality. The most recent version of the Apple OS took four months longer to develop than the previous version.[22] Microsoft's Vista operating system has also been plagued with very long delays.

3. Congestion and Search Optimization

Several design issues tend to limit the size of two-sided platforms. Physical platforms such as trading floors, singles clubs, auction houses, and shopping malls help customers search for and consummate mutually advantageous exchanges. At a given size expanding the number of customers on the platform can result in congestion that increases search and transaction costs.[23] It may be possible to reduce congestion by increasing the size of the physical platform, but that in turn may increase search costs. Indeed, to optimize searching for partners, two-sided platforms may find that it is best to limit the size of the platform and prescreen the customers on both sides to increase the probability of a match. One might argue that singles-type clubs do this explicitly (deciding who can get into an "exclusive" club) or implicitly (compare church-oriented singles groups and Club Med resorts). We will return to this subject below in discussing platform differentiation. Congestion may arise on one side alone. For example, increasing the volume of advertising in a newspaper may not only crowd out the content that attracts the readers but also result in a cacophony of messages that reduces the effectiveness of any particular advertisement.

4. Platform Differentiation and Multi-Homing

Platforms can differentiate themselves from each other by choosing particular levels of quality (what is known as "vertical differentiation") with consumers choosing the higher or lower quality of platform depending on the income and relative demand for quality. There are, for example, upscale and downscale malls. Platforms can also differentiate themselves from each other by choosing particular features and prices that appeal to particular groups of customers (what is known as "horizontal differentiation"). Thus there are numerous advertising supported magazines that appeal to particular segments of readers and advertisers (e.g., Cape Cod Bride or Fly Fisherman).

Horizontal differentiation can result in customers choosing to join and use several platforms—a phenomenon that Rochet and Tirole have called "multi-homing". Customers find certain features of different competing platforms attractive and therefore rely on several. Payment cards are an

[22] For Apple OS release dates, *see* Jason Snell, *Jaguar unleashed: Mac OS X 10.2 Arrives*, Macworld, Sept. 1, 2002; Sarah Stokely, *Apple Sets Panther Release Date*, IDG Data, Oct. 10, 2003.; and, Steven Musil, *This Week in Tiger: Apple releases Mac OS X 10.4*, CNET NEWS, Apr. 29, 2005.

[23] For a general discussion on matching, search, and congestion see, for example, Robert Shimer & Lones Smith, *Matching, Search, and Heterogeneity*, 1 Advances In Macroeconomics (2001) and Mark Rysman, *Competition Between Networks: A Study of the Market for Yellow Pages,* 71 Rev. Econ. Studies 483 (2004b).

example of multi-homing on both sides. Most merchants accept credit and debit cards from several systems, including ones that have relatively small shares of cardholders. Many cardholders carry multiple cards, although they may tend to use a favorite one most often.[24] Advertising supported media also has multi-homing on both sides—advertisers and viewers rely on many differentiated platforms. Other two-sided platforms have multi-homing only on one side. Most end-users rely on a single software platform for their personal computers, for instance, while many developers write for several platforms.

B. Empirical Evidence on Two-Sided Industry Structure

It is possible to see some regularities across industries in which two-sided platforms appear to be the dominant form of organization. Table 1 above and Table 3 reveal several features:

- It is relatively uncommon for industries based on two-sided platforms to be monopolies or near monopolies. Some industries based on two-sided platforms have several large differentiated platforms, while others have many small platforms that are differentiated by location as well as along other dimensions.
- Multi-homing on at least one side is common. Horizontal product differentiation tends to be the norm.
- Asymmetric pricing is relatively common. Many two-sided platforms appear to obtain the preponderance of their operating profits (revenues minus direct costs) from one side. A nontrivial portion of two-sided platforms appear to charge prices that are below marginal cost or below zero.

V. OVERVIEW OF ANTITRUST CASES INVOLVING TWO-SIDED MARKETS

Many antitrust cases have involved two-sided platforms. A few—including several important ones—seem to have touched on two-sided issues before economists began to address them formally. And some are based on analyses of markets and practices that, putting aside whether they led to the correct outcome or not, are analytically wrong from the perspective of the two-sided literature.

Table 4 presents an overview of antitrust cases in the European Community and the United States that concern two-sided platforms. We have not done a systematic review of cases but have rather listed cases that have had a high profile in these jurisdictions with which we are gener-

[24] Mark Rysman, An Empirical Analysis of Payment Card Usage (Boston University Department of Economics, Working Paper, 2004).

TABLE 3 Presence of multi-homing and largest competitor share of selected two-sided platforms

Multi-Sided Platform	Sides	Presence of Multi-homing	Largest Competitor Share in the United States
Residential Property Brokerage	Buyer Seller	**Uncommon:** Multi-homing may be unnecessary, since a multiple listing service allows the listed property to be seen by all member agencies' customers and agents.	Fifty largest firms have a 23% share. (2002)
Securities Brokerage	Buyer Seller	**Common:** The average securities brokerage client has accounts at three firms. Note that clients can be either buyers or sellers or both.	Four largest firms accounted for 37% of in securities brokerage and 16% in financial portfolio management. (2002)
Newspapers and Magazines	Reader Advertiser	**Common:** In 1996, the average number of magazine issues read per person per month was 12.3. **Also common for advertisers:** for example, AT&T Wireless advertised in the New York Times, The Wall Street Journal, and Chicago Tribune, among many other newspapers, on Aug. 26, 2003.	Wall Street Journal had a 28% share of the five largest newspapers. (2001)
Network Television	Viewer Advertiser	**Common:** For example, viewers in Boston, Chicago, Los Angeles, and Houston, among other major metropolitan areas, have access to at least four main network television channels: ABC, CBS, FOX, and NBC. Also common for advertisers: for example, Sprint places television advertisements on ABC, CBS, FOX, and NBC.	U.S. law forbids broadcasters from owning TV stations reaching more than 35% of the nation's television audience.
Operating System	End User Application Developer	**Uncommon for users:** Individuals typically use only one operating system. Common for developers: As noted earlier, the number of developers that develop for various operating systems indicates that developers engage in significant multi-homing.	Microsoft has a 96% share of revenue of client operating systems. (2004)
Video Game Console	Game Player Game Developer	**Varies for players:** The average household (that owns at least one console) owns 1.4 consoles. **Common for developers:** For example, in 2003, Electronic Arts, a game developer, developed for the Nintendo, Microsoft, and Sony platforms.	Sony PS1 and PS2 had a 63% share of console shipments in North America. (2003)
Payment Card	Cardholder Merchant	**Common:** Most American Express cardholders also carry at least one Visa or MasterCard. In addition, American Express cardholders can use Visa and MasterCard at almost all places that take American Express.	The Visa system had a 45% share of all credit, charge, and debit purchase volume. (2004)

Source: Adapted from David S. Evans, *The Antitrust Economics of Multi-Sided Platform Markets*, 20 Yale J. ON Reg. 325 (2003). Industry share data from United States Census Bureau, 2002 Economic Census, available at http://www.census.gov/econ/census02/guide/INDSUMM.HTM; "Top 20 U.S. Daily Newspapers by Circulation," Newspaper Association of America (2001), at http://www.naa.org/info/facts01/18_top20circ/index.html (accessed Feb. 21, 2007); Stephen Labaton, *U.S. Backs Off Rules for Big Media*, NY TIMES, Jan. 28, 2005; A. Gillen & D. Kusnetzky, *Worldwide Client and Server Operating Environments 2004-2008* Forecast, IDC Market Analysis, No. 32452 (Dec. 2004); Schelley Olhava, *Worldwide Videogame Hardware and Software 2004-2008* Forecast and Analysis, IDC Market Analysis, No. 31260 (May 2004); The Nilson Report, No. 828 (Feb. 2005); The Nilson Report, No. 833 (May 2005).

TABLE 4[25] Summary of leading cases by two-sided platform type

	Case	Case Type
Media	Times Picayune	Monopolization
	Magill	Refusal to supply
	BT Yellow Pages	Monopolization
	Lorain Journal	Exclusive dealing
Exchanges	Sotheby's–Christies	Cartel
	Marsh McLennan	Cartel
	Stock Exchanges	Merger
	Mobile operators	Excessive Pricing

	Case	Case Type
Transaction Systems	NaBanco	Cartel
	Wal–Mart	Tying
Software Platforms	Microsoft–Browser	Monopolization, Tying
	Microsoft–Media Player	Tying
	Nintendo	Exclusivity

ally familiar.[26] The cases span all of the major categories of two-sided platforms and involve the spectrum of competition policy issues. This section summarizes some key issues that arose in several of these cases.

A. *NaBanco*

In *NaBanco v. Visa*, the federal district court and the U.S. Court of Appeals for the Eleventh Circuit recognized several of the key features of what have become known as two-sided platforms. Visa was (and is) a cooperative of banks that issued cards and acquired those card transactions from merchants. It established a rule for governing the situation in which an individual whose card was

[25] United States v. Times-Picayune Publishing Co., 345 U.S. 594 (1953); Joined Cases C-241/91 P and C-242/91 P, RTE, BBC, and ITP v. Commission of the European Communities (*Magill*), 1995 E.C.R. I-00743 (Apr. 6, 1995); U.K. Competition Commission, Classified Directory Advertising Services (1996); U.K. Office of Fair Trading, Classified Directory Advertising Services: Review of Undertakings Given by BT to the Secretary of State in July 1996 (2001); United States v. Lorain Journal Co., 342 U.S. 143 (1951); United States v. Taubman, 297 F.3d 161 (2d Cir. 2002); State of New York v. Marsh & McLennan Companies, Inc., et al., Complaint filed October 14, 2004, Index No. 04-403342; U.K. Competition Commission, A Report on the Proposed Acquisition of London Stock Exchange PLC by Deutsche Börse AG or Euronext NV (2005); U.S. Department of Justice, Department of Justice Antitrust Division Statement on the Closing of its Two Stock Exchange Investigations (Nov. 16, 2005); U.K. Office of Communications, Wholesale Mobile Voice Call Termination (Jun. 1, 2004); National Bancard Corp. v. Visa U.S.A., Inc., 779 F.2d 592, 602 (11th Cir. 1986); In re Visa Check/MasterMoney Antitrust Litigation, 192 F.R.D. 68 (E.D.N.Y 2000); United States v. Microsoft, 87 F. Supp. 2d 30 (D.D.C. 2000); Commission of the European Communities v. Microsoft, Case COMP/C-3/37.792/Microsoft; Atari Games Corp. v. Nintendo, 975 F.2d 832 (Fed. Cir. 1992).

[26] *See* J. Wotton's article in this issue (John Wotton, *Are Media Markets Analyzed as Two-Sided Markets?,* (3)1 Competition Pol'y Int'l 237–47 (2007)).

issued by bank A paid with that card at a merchant acquired by bank B, where A and B are different banks. Although those banks could have a bilateral agreement, Visa established a default rule that among other things determined the allocation of the profits and risks of the transaction. This rule provided that given the various allocations of risks and costs that the bank that acquired the transaction (B) had to pay the bank (A) that issued the card a percent of the transaction amount; this percent is known as the interchange fee, and it was initially set at 1.95 percent.

NaBanco argued that the interchange fee violated Section 1 of the Sherman Act because it was a price set collectively by competitors. Visa argued that unlike classic price-fixing, the ability to set an interchange fee was a mechanism to allocate costs between the issuing and acquiring sides of the business and enhanced output by, among other things, limiting opportunistic behavior by individual members and avoiding the chaos of bilateral negotiations among thousands of member banks. The Eleventh Circuit concluded:

> "Another justification for evaluating the [interchange fee] under the rule of reason is because it is a potentially efficiency creating agreement among members of a joint enterprise. There are two possible sources of revenue in the VISA system: the cardholders and the merchants. As a practical matter, the card-issuing and merchant-signing members have a mutually dependent relationship. If the revenue produced by the cardholders is insufficient to cover the card-issuers' costs, the service will be cut back or eliminated. The result would be a decline in card use and a concomitant reduction in merchant-signing banks' revenues. In short, the cardholder cannot use his card unless the merchant accepts it and the merchant cannot accept the card unless the cardholder uses one. Hence, the [interchange fee] accompanies "the coordination of other productive or distributive efforts of the parties" that is 'capable of increasing the integration's efficiency and no broader than required for that purpose.'"[27]

Professor William Baxter worked for Visa on this matter. His 1983 article in the *Journal of Law and Economics* presented many of the key concepts of two-sided markets within the context of the determination of interchange fees.[28] The modern literature now recognizes that the interchange fee is at least partly a device for determining the pricing structure for the card system.[29] Some regulators and antitrust authorities, while recognizing the two-sided nature of the business, have argued in recent years that the interchange fee is set at a level that encourages the overuse of cards.

---

[27] National Bancard Corp. v. Visa U.S.A., Inc., 779 F.2d 592, 602 (11th Cir. 1986).

[28] Baxter, *supra* note 2.

[29] *See, e.g.,* Richard Schmalensee, *Payment Systems and Interchange Fees*, 50 J. Indus. Econ. 103 (2002); Jean-Charles Rochet & Jean Tirole, *Cooperation among Competitors: Some Economics of Credit Card* Associations, 33 Rand J. Econ. 549 (2002); See Rochet & Tirole, supra note 2; See Wright, supra note 1; David S. Evans & Richard Schmalensee, Paying with Plastic: The Digital Revolution in Buying and Borrowing (MIT Press 2005); and Evans & Schmalensee (2005), supra note 13.

B. Stock Exchange Mergers

In recent years, stock exchanges have increasingly looked to merge with each other. In December 2004, Euronext and Deutsche Börse, respectively the second and third largest stock exchanges in Europe by value of trading, made bids to take over the London Stock Exchange, the largest stock exchange in Europe. Both bids were referred to the U.K.'s Competition Commission for investigation under U.K. competition law—they did not qualify for investigation by the European Commission under EU law. In its report, the Competition Commission expressed concerns about the ownership of clearing services by the Euronext or Deutsche Börse that was likely to result post merger. It was believed that ownership of clearing services by the London Stock Exchange's parent company would act as a barrier to potential competitor exchanges to the London Stock Exchange that needed access to same clearing service to be competitive. Both Euronext and Deutsche Börse made commitments that satisfied the concerns of the Competition Commission but as a result of business rather than regulatory reasons, neither deal went through.

In the United States, in 2005 the New York Stock Exchange agreed to merge with Archipelago, an electronic stock exchange, and the NASDAQ Stock Exchange agreed to merge with Instinet, also an electronic stock exchange. The U.S. Department of Justice approved both mergers, in part because it believed that there were no likely anticompetitive effects given the planned and likely entry of other firms. In 2006, the New York Stock Exchange and Euronext announced they had agreed to merge. As of this writing, the transaction has recently received antitrust and regulatory approval in the United States and Europe, but has not yet been consummated.

Stock and other exchanges exhibit significant network effects. Fundamentally, more trading activity on the part of providers and consumers of liquidity tends to reduce spreads between bid and ask prices and to make markets more liquid, so that large blocks of stocks, options, or commodities can be bought or sold rapidly without a price penalty. And, of course, smaller bid-ask spreads and more liquidity tend to attract more trading. The more investors that come to a market, the more attractive that market becomes to liquidity providers, and the more liquidity providers are present, the more attractive the market is to investors.[30]

Traditionally, stock exchanges have tended to be local monopolies, due in large part to these network effects, to regulations that restricted cross-border trading and, historically in the United States, to communications costs that created a niche for regional exchanges like the Boston Stock Exchange. As these restrictions have been relaxed and communications costs have fallen, competition has increased generally, and many exchanges have abandoned their traditional non-profit, cooperative structures and become for-profit firms. In the United States, regional stock exchanges have had trouble competing with the NYSE, but competition between the NYSE and NASDAQ has intensified. There are now six competitive equity options exchanges in the United States; they are linked electronically so that investors are guaranteed the best available price, and the largest market shares

---

[30] *See* Friess & Greenaway, *supra* note 9.

hover below 40 percent. Stock exchanges have been ordered to provide such linkage; this is expected to happen in the first half of 2007 and may have a major effect on the competitive landscape.

In Europe, on the other hand, there has thus far been very little direct competition between the London Stock Exchange and other European exchanges, such as Euronext and Deutsche Börse. One key question in mergers between stock exchanges is whether network effects will continue to limit the scope for competition or whether falling communications costs and the computerization of the securities business will make global competition—of one sort or another—inevitable.

C. Microsoft Media Player

The European Commission found that Microsoft had abused a dominant position in operating systems by including media player technologies in Windows.[31] It argued that there were indirect network effects between the use of media players and the provision of content. If more people have a particular media player, content providers will tend to encode content in that format. If more content is available in the format for a particular media player, users will tend to use that media player. The Commission argued that content providers would standardize on Windows Media Player because this player was available on most personal computers, which of course included Windows. In effect, the Commission argued that the existence of network effects would result in the "media player market" tipping to Windows Media Player.[32]

For its part Microsoft has agreed that there are indirect network effects but that the existence of such effects is not sufficient to tip a market to a single platform. In particular, it has argued that media players are horizontally differentiated products and that most content providers and many users engage in multi-homing. Who is right on this score depends on factual disputes between the Commission and Microsoft that we do not consider here.

D. *Magill*

Magill is a leading EC case involving the compulsory licensing of intellectual property. What makes it interesting from a two-sided standpoint is that it involved several interlinked two-sided platforms. The defendants in the case were three television networks (RTE, BBC, and ITV) whose broadcasts were received in Ireland. RTE and ITV were two-sided platforms, receiving revenues from advertisers. RTE was also supported by licenses paid by consumers for having television sets. The BBC received similar revenues from licenses for television sets in the United Kingdom (but not

---

[31] For contrary views on this case, *see* Maurits Dolmans & Thomas Graf, *Analysis of Tying Under Article 82 EC: The European Commission's Microsoft Decision in Perspective*, 27 World Competition 225 (2004). See also David S. Evans & A. Jorge Padilla, *Tying Under Article 82 EC and the Microsoft Decision: A Comment on Dolmans and Graf*, 27 World Competition 503 (2004).

[32] CFI Order of Dec. 22, 2004, Case T-201/04 R 2, Microsoft Corporation v. Commission, at paras. 365 and 388, *available at* http://curia.eu.int/en/content/juris/index_form.htm.

Ireland). The BBC did not allow advertising and was not a two-sided platform. All three networks published an advertising-supported television guide that contained their own weekly listings; these were two-sided platforms. In addition they each provided their daily listings to newspapers—other two-sided platforms—that combined the listings.

Magill TV Guide (Magill) wanted to publish a weekly advertising-supported guide that contained the listings of the three networks. The networks complained that this violated their copyrights. The Commission and ultimately the EC courts concluded that there would be a market—in the antitrust sense—for a weekly television guide and that the refusal to supply the copyrighted information prevented the emergence of the weekly guide product. As it turns out, the weekly newspapers were the main beneficiaries of this decision since they started weekly television supplements included in the Sunday newspapers. Magill never made a successful go of it.

We will return to these issues when we discuss the analysis of market definition and market power. The key point is that the analysis by all the parties (including the television networks) ignores a key side of the two-sided industry here—the advertisers who were the likely source of much of the revenue and profits—as well as the link between the guides and the television business.

VI. ANTITRUST IMPLICATIONS OF TWO-SIDED PLATFORM ECONOMICS

Whether the economics of two-sided platforms can assist in determining whether a merger or business practice is anticompetitive is, like many aspects of economics, an empirical question. As with market power generally two-sidedness is a matter of degree. Sometimes the two-sided nature of the business is critical for the analysis. Other times it is an interesting aspect of the industry that should be thought about but is not ultimately determinative. And still other times an industry may have two-sided aspects that are too insubstantial to matter.

A. Market Definition and Market Power

The analysis of market power, and the associated issue of the definition of the relevant market are typically a central component of antitrust cases, although the reasons for this vary somewhat across antitrust matters. In most cases it is crucial to determine whether the defendants have or could obtain significant market power and thus, by definition, maintain or raise prices above the competitive level. The determination of whether a firm or group of firms has market power can also be important because entities that have significant market power are more likely to have the ability and incentive to engage in business practices that could foreclose competition. Moreover, entities that obtain significant market power as a result of a business practice may be able to recoup costs they incur from investing in anticompetitive activities such as predatory pricing and vertical foreclosure. Business practices engaged in by entities that either lack market power or are unlikely to acquire it are often presumed benign (except of course for naked price-fixing and related cartel practices).

The economics of two-sided platforms provides several insights into analysis of market power.

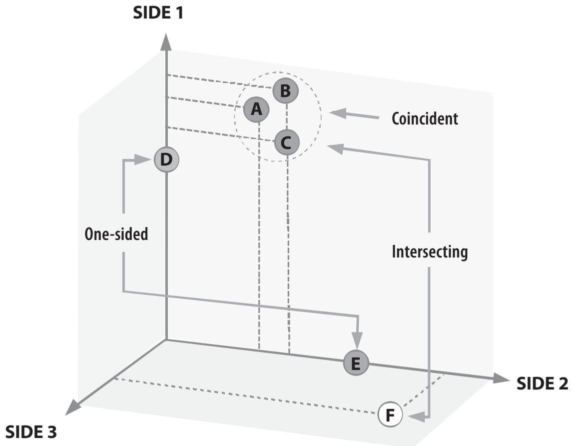
1. The link between the customers on the two-sides affects the price elasticity of demand and thus the extent to which a price increase on either side is profitable. It therefore necessarily limits market power all else equal. Consider two sides ***A*** and ***B.*** An increase in the price to side ***A*** reduces the number of customers on side ***A*** and therefore reduces the value that customers on side ***B*** receive from the platform. That in turn reduces the price that side ***B*** will pay and the number of customers on side ***B.*** The reduction in the number of customers on side ***B*** in turn reduces the demand on side ***A*** and thus the price that customers on side ***A*** will pay. These positive feedback effects may take some time to work themselves out, but, as we demonstrated above, even if, say, customers on side ***A*** are not very sensitive to price, all else (including the behavior of those in side **B)** equal, demand from side ***A*** may nonetheless end up being very price-sensitive indeed when these feedback effects work themselves out.

2. For two-sided platforms it can be important to recognize that competition on both sides of a transaction can limit profits. Suppose in a market without multi-homing that there is limited competition on side ***A*** because customers cannot easily switch between vendors of that side, but there is intense competition on side ***B*** because customers can and do switch between vendors based on price and quality. Then if competitors on side ***B*** cannot differentiate their products and otherwise compete on an equal footing, the ability to raise prices on side ***A*** will not lead to an increase in profits. Any additional profits on side ***A*** will be competed away on side ***B.*** This is different from a simple multi-product setting, since the platform cannot stop serving side *B* without leaving the business entirely. This point is especially relevant for assessing incentives and recoupment. It is also worth noting that the possibility of multi-homing on side ***B*** will permit positive profits, since it reduces the intensity of competition.

3. Price equals marginal cost (or average variable cost) on a particular side is not a relevant economic benchmark for two-sided platforms for evaluating either market power, claims of predatory pricing, or excessive pricing under EC law. As we saw above, the non-predatory, profit-maximizing price on each side is a complex function of the elasticities of demand on both sides, indirect network effects, and marginal costs on both sides. Thus it is incorrect to conclude, as a matter of economics, that deviations between price and marginal cost on one side provide any indication of pricing to exploit market power or to drive out competition.[33]

[33] For the two-sided platform as a whole, a formula similar to the standard Lerner index emerges in the Rochet-Tirole model. This is not a general result, and it thus suggests that the overall price-cost margin is somewhat less relevant than in single-sided businesses for evaluating overall market power.

FIGURE1 Types of differentiated platform competition



	Coincident			One-sided		Intersecting
	A	B	C	D	E	F
SIDE 1	X	X	X	X		
SIDE 2	X	X	X		X	X
SIDE 3						X

The constraints on market power that result from interlinked demand also affect market definition. Market definition assists in understanding constraints on business behavior and assessing the contours of competition that are relevant for evaluating a practice. In some cases, the fact that a business can be thought of as two-sided may be irrelevant. That could happen either because the indirect network effects though present are small or because nothing in the analysis of the practices really hinges on the linkages between the demands of participating groups. In other cases, the fact that a business is two-sided will prove important both by identifying the real dimensions of competition and focusing on sources of constraints.[34]

Figure 1 shows potential sources of competitive constraints for a two-sided platform denoted by ***A***. It faces competition of some degree from other differentiated two-sided platforms that serve the same customer groups (e.g., the newspapers in a city). It also faces competition from single-sided businesses that provide competitive services to one side only (e.g., billboards). And it faces

[34] *See* David S. Evans & Michael Noel, *Defining Antitrust Markets When Firms Operate Multi-Sided Platforms*, 3 Colum. Bus. L. Rev. 667 (2005).

competition from other two-sided platforms that provide a product that competes mainly with one side but not the other (e.g., advertising-supported television). Again, the existence of these constraints does not mean they are important, only that they need to be looked at.

B. Coordinated Practices

The key insight of the economics of two-sided platforms in the oligopoly context is that to be successful cartels may need to coordinate on both sides. Consider the situation in which there are several competing two-sided platforms. If they agree to fix prices on one side only the cartel members will tend to compete the supracompetitive profits away on the other side. This observation has two corollaries. The first is that it is harder to form an effective cartel in an industry with two-sided platforms than in single-sided industries, all else equal. The cartel requires more agreements and monitoring because of the additional side. The second is that if an authority finds evidence of a price fix on one side it should probably look carefully for evidence on the other side. This was relevant, as we note above, in the price-fixing case involving Sotheby's and Christie's.

The economics of two-sided platforms is also relevant for evaluating the practices of cooperatives and joint ventures as we saw from the discussion of the *NaBanco* case. Payment card systems, financial exchanges, and music collecting societies are examples of two-sided platforms that are sometimes organized as not-for-profit cooperatives. The two-sided platforms adopt various rules and regulations for the members and take charge of certain centralized functions. The economics of two-sided platforms is useful for assessing whether there is an efficiency rationale behind an agreement over prices. In *NaBanco*, as we noted, the court found that the collective setting of the interchange fee helped balance the demands between cardholders and merchants (it helped internalize an externality) and eliminated the need for bilateral negotiations (it reduced the transactions cost of internalizing the externality).

C. Unilateral Practices

In trying to assess whether unilateral practices are anticompetitive the special economic features of two-sided platforms need to be considered.

1. Predatory and Excessive Pricing

Our review of pricing showed that a robust conclusion of the economics literature is that profit-maximizing two-sided platforms may find that it is profitable overall to price the product offered on one side below average variable cost, below marginal cost, or even below zero. The empirical evidence indicates that such below-cost pricing is common, occurs in stable market equilibrium, and is therefore not designed mainly for the purpose of foreclosing competition. Therefore, any presumption that below-cost pricing by two-sided platforms is anticompetitive is simply not valid. Of course, it is certainly possible for two-sided platforms to engage in predatory pricing by setting its price on one side so low as to deny other platforms access to this side of the market. It is also possible for a two-sided platform to engage in two-sided predatory pricing, charging below cost overall on both sides with the purpose of foreclosing competitors. Cost-based tests make some sense in the latter case, but it is hard to see how they could be used to analyze an allegation of one-sided predation.

Under Article 82 of the EC Treaty a dominant firm can be found to have made an abuse by charging "unfair purchase or selling prices." Just as a below-cost price on one side can emerge in long-run market equilibrium so can an above-cost price on the other side. Indeed, such below-cost/above-cost prices will come together. This issue has come up in a series of cases in Europe in which regulatory authorities have found mobile telephone operators to have charged fixed-line carriers excessive prices for terminating calls on their networks; the authorities recognize that the profits from these excessive prices are competed away in part through low prices for handsets and call origination. Indeed, the U.K.'s Office of Communication (OfCom) recognized that mobile telephone platforms were highly competitive (on the mobile subscriber side at least) and did not overall earn supracompetitive returns.[35] Although they did not accept that this was a two-sided business, and did not apply two-sided analysis, OfCom did provide an "indirect network externality" kicker to the regulated price it imposed on the mobile termination side.[36]

2. Tying

Under a rule of reason analysis[37] the economics of two-sided platforms can provide an explanation for certain tying practices that seem to reduce consumer choice and harm consumers. As we discussed above, the platform provider designs the platform—including the constellation of services and features—to harness internalized externalities, minimize transactions costs between the customers and both sides, and maximize the overall value of the platform. As part of harnessing externalities this platform provider wants to increase positive indirect network effects while limiting negative

---

[35] *See, e.g.,* U.K. Office of Telecommunications, Discontinuing Regulation: Mobile Access and Call Origination Market §1.2 (2003), *available at* http://ofcom.org.uk/static/archive/oftel/publications/eu_directives/2003/discon1103.pdf ("no mobile network operator, either individually or in combination with one or more other mobile network operators, has [significant market power] in that market."). No provider has a share exceeding 28 percent. *See, e.g.*, Economist Intelligence Unit, United Kingdom: Telecoms and Technology Background (2005).

[36] U.K. Office Of Communications, Wholesale Mobile Voice Call Termination 163-72 (2004), *available at* http://www.ofcom.org.uk/consult/condocs/mobile_call_termination/wmvct/wmvct.pdf. *See* Armstrong, *supra* note 8.

[37] Economists and legal scholars generally agree that tying should be considered under a rule of reason analysis rather than a *per se* test. That is not the state of the law in the United States or the European Community, both of whose highest courts have adopted something closer to a *per se* test of liability. However, both courts admit that efficiencies can at least play a limited role in the analysis (in the United States through the separate product test and in the European Union through the possibility of "objective justification" of the practice).

indirect network effects. As a consequence, the two-sided platform may impose requirements on side ***A*** that do not benefit them directly and which customers on that side might even reject after comparing private benefits and costs. But such requirements may benefit side ***B***. And if the demand increases on side ***B***, these requirements may increase the value placed on the platform on side ***A***—and in fact could increase value so much that the feature provides a net benefit to side ***A***.[38]

Under a rule of reason analysis the economics of two-sided platforms can provide an explanation for certain tying practices that seem to reduce consumer choice and harm consumers.

The honor-all-cards rule for payment cards is a possible example. Card systems generally require that merchants that agree to take the system's branded cards agree to take all branded cards that are presented by shoppers. Thus, merchants that have a contract to take American Express (Amex) cards cannot decide to take payment by Amex corporate cards but not Amex personal cards, or to take payment from visibly wealthy travelers but not from locals. For at least some merchants the private cost of this requirement outweighs its benefits (generally we would expect that merchants would privately want a choice to take whatever card they wanted).[39] However, this rule makes the system's branded card more valuable to its cardholders, who have the assurance that their card will be accepted for payment at merchants that display the system's acceptance mark. By increasing the number of cardholders it makes the card a more valuable payment device for merchants to accept.[40]

3. Exclusive Dealing

The potential for profits on the other side provides a possible incentive for exclusive contracts in two-sided platforms. One of the main Chicago School observations about exclusive contracts is that a consumer is always free not to agree to exclusivity. The conclusion is that exclusivity in contracts must reflect consumers' judgment that the benefits (lower prices or efficiencies) outweigh the costs of only dealing with one firm. For two-sided platform businesses, it is at least possible that there is an externality; exclusive contracts on one side might help a platform gain market power on other sides. The consumers agreeing to the exclusive contracts on one side might, at least in the short run, gain from or be indifferent to exclusivity, but they may not take into account the costs to consumers on the other sides from decreased platform competition. Some recent work suggests

---

[38] *See* Rochet & Tirole (2005), *supra* note 6.

[39] For a discussion of this issue, *see* Robert E. Litan & Alex J. Pollock, The Future of Charge Card Networks (AEI-Brookings Joint Center for Regulatory Studies, Working Paper, 2006).

[40] A class of merchants claimed that Visa and MasterCard had illegally tied by requiring merchants that accepted their credit cards to also accept their debit cards. The card associations agreed to end this practice after a federal district court judge applied the *per se* tying test and ruled that the associates failed several prongs of this test as a matter of law. In re Visa Check/MasterMoney Antitrust Litigation, 192 F.R.D. 68 (E.D.N.Y 2000). American Express has been sued by a class of merchants for illegally tying its corporate and personal cards. *See* Lavonne Kuykendall, *Merchants Suing Amex Add Citi, MBNA as Defendants*, 170 Am. Banker (2005).

that it is at least theoretically possible for a two-sided platform to use exclusive contracts to exclude competitors, although the welfare consequences of these contracts are not clearly harmful.[41]

As with exclusivity in one-sided markets, however, this can only be a concern if one firm has exclusivity over most or all of the market and if the exclusivity is persistent and durable. For example, consumers on the nonexclusive side could respond by moving to a competing platform, thus exerting pressure on consumers on the exclusive side to end exclusivity. Moreover, in markets with significant buyer concentration, the buyers would be reluctant to agree to exclusivity if there is some expectation that it will lead to dominance by that platform, as that will likely result in higher prices in the future for all sides. As with one-sided markets, one needs to consider whether the efficiencies from exclusive contracts—for example, in helping to create a platform that might not otherwise exist for the benefit of consumers—offset possible costs from reducing competition.

VII. QUALIFICATIONS AND CONCLUSIONS

The indirect network effects between customer groups served by a single business are strong in many important industries. Businesses in these industries operate two-sided platforms. The economics of two-sided platforms provides insights into how these businesses and industries behave that are relevant for competition analysis including market definition, coordinated practices, unilateral practices, and the evaluation of efficiencies. The economic literature provides robust results—that is, ones that are not dependent on only fragile assumptions—that can assist in this analysis. These results include the consequences of interlinked demand between customer sides for prices; prices do not, contrary to the standard model, have a tight relationship with cost.

As with almost any application of economics to policy several cautions are prudent. First, many of the theoretical results in the literature to date are, like those in other areas of industrial organization, based on quite abstract models of how industries operate and special assumptions of demand and cost. Second, to date there has been little rigorous empirical research on two-sided platforms or competition among them. Third, the theoretical and empirical work to date suggests that how two-sided businesses work is highly dependent on the specific institutions and technologies of an industry. One must be careful generalizing.

[41] *See* Mark Armstrong & Julian Wright, Two-Sided Markets, *Competitive Bottlenecks and Exclusive Contracts*, Econ. Theory (forthcoming 2006).

CHAPTER

2

Some Empirical Aspects of Multi-Sided Platform Industries

David S. Evans

ABSTRACT

Multi-sided platform markets have two or more different groups of customers that businesses have to get and keep on board to succeed. These industries range from dating clubs (men and women), to video game consoles (game developers and users), to payment cards (cardholders and merchants), to operating system software (application developers and users). They include some of the most important industries in the economy. A survey of businesses in these industries shows that multi-sided platform businesses devise entry strategies to get multiple sides of the market on board and devise pricing, product, and other competitive strategies to keep multiple customer groups on a common platform that internalizes externalities across members of these groups.

I. INTRODUCTION

Multi-sided platforms coordinate the demand of distinct groups of customers who need each other in some way. Dating clubs, for example, enable men and women to meet each other; yellow pages provide a way for buyers and sellers to find each other; and computer operating system vendors provide software that applications users, applications developers, and hardware providers can use together. When devising pricing and investment strategies, multi-sided platforms must account for interactions between the demands of multiple groups of customers. In theory, the optimal price to customers on one side of the platform is not based on a markup formula such as that given by the Lerner condition, and price does not track marginal cost. Competition among platforms takes place when seemingly distinct customer groups are connected through interdependent demand and a platform that, acting as an intermediary, internalizes the resulting indirect network externalities (Rochet and Tirole, 2003). Platforms are central to many key industries including computer games, information technology, many internet-based industries, media, mobile telephony and other telecommunications industries, and payment systems.

This chapter provides an empirical survey of entry, pricing and other strategies in platform industries. It provides background for the emerging theoretical (Rochet and Tirole, 2003; Julien, 2001; Armstrong, 2002; and Parker and Van Alstyne, 2002); empirical (Rysman, 2002), and policy (Evans, 2003) literature on multi-sided platform markets. Although it does not provide empirical tests of either the key assumptions of this literature or its implications, it confirms that multi-sided platforms are an important, and until recently unrecognized, part of the industrial organization landscape, and that this new area of economic research has potentially rich empirical implications and relevance. This chapter is based, and reports early results, of a series of detailed case studies of multi-sided platform industries I have been conducting.

Section 2 summarizes the conditions under which a multi-sided platform may emerge and the main theoretical findings of the literature to date. It also provides a brief overview of the three major kinds of platforms. Section 3 reviews some common business practices followed in multi-sided platform industries studied thus far. I then turn to two more detailed case studies. Section 4 reviews the entry of Diners Club in the payment card industry and summarizes the pricing strategies that continue to this day. Section 5 examines the entry of the Palm operating system for personal digital assistants. Section 6 makes some brief concluding remarks.

II. A BRIEF REVIEW OF THE ECONOMICS OF MULTI-SIDED PLATFORM MARKETS

There is an opportunity for a platform to increase social surplus when three necessary conditions are true: (1) there are distinct groups of customers; (2) a member of one group benefits from having his demand coordinated with one or more members of another group; and (3) an intermediary can facilitate that coordination more efficiently than bilateral relationships between the members of the group.[1] As an empirical matter, indirect network effects generally accompany condition (2) and intimately shape the business strategies in these industries along side the multi-sidedness.

1. There are two or more distinct groups of customers. These customers may be quite different from each other, such as the men and women for a dating platform or retailers and customers for a shopping mall. Alternatively, these customers may be different only for the purpose of the transaction at hand – eBay users are sometimes buyers, sometimes sellers; mobile phone users are sometimes callers, sometimes receivers.

2. There are externalities associated with customers A and B becoming connected or coordinated in some fashion. A cardholder benefits when a merchant takes his card for payment; a merchant benefits when a cardholder has a form of payment he takes. The presence of indirect network effects seems to be an empirically important explanation for the emergence

[1] See Rochet and Tirole (2002), Rochet and Tirole (2003), Rochet (2003), Armstrong (2002), and Parker and Van Alstyne (2002).

of a platform although not necessary as a matter of theory. Sellers of PEZ dispensers value exchanges that have more people who would like to buy PEZ dispensers. (See discussion of eBay below.)

3. An intermediary can internalize the externalities created by one group for the other group. Obviously, if the members of group A and group B could enter into bilateral transactions they would be able to internalize the indirect externalities under the second condition (2). However, in practice information and transaction costs and free-riding problems make it difficult for members of distinct customer groups to internalize the externalities on their own.[2] You can look for your new sweetheart by strolling around the Boston Public Garden; the Yahoo! Personals are less romantic but perhaps more efficient. The intermediary does not have to be a business in the usual sense. Cooperatives have emerged in payment cards (Visa International) and software (Linux). Governments sometimes act as the intermediary – currency is an example.

Several articles have examined the economics of price determination in multi-sided platform markets. A key finding is that optimal prices for the multiple customer groups must align – or balance – the demand among these groups – and indeed the emergence of a pricing structure as well as a pricing level is the defining characteristic of such industries (Rochet and Tirole, 2002). Optimal prices are not proportional to marginal costs as is the case with the familiar Lerner conditions[3] or its multi-product variants.[4] Indeed, it is possible that the optimal price for one side will be less than the marginal cost for that side (Parker and Van Alstyne, 2002). (The assignment of costs to one side or another may not be well defined either. When it is necessary to get both sides together for a platform product to exist – that is for either customer to have anything to purchase – one may not be able to say that one side or another "caused" a cost.) Platform businesses may tend to skew prices towards one side or another depending upon the magnitude of the indirect network externalities resulting from that side. If side A generates a much greater degree of externalities for side B than side B does for side A, side A may tend to get a lower price (Parker and Van Alstyne, 2002). As in Ramsey-type models of multi-product pricing, in which firms are pricing in part to recover common costs of production, one side may end up contributing more to common costs than another side. However, the economic reasons for this are different in multi-sided than in single-sided markets.[5] Note that the theoretical literature is based on quite rarefied assumptions and has thus far focused on static pricing issues.

---

[2] In unpublished work Jean Tirole points out that a necessary condition for multi-sided platforms to arise is that the Coase Theorem does not apply.

[3] The Lerner condition was first stated in Lerner (1934).

[4] See generally Baumol, Panzar and Willig (1982).

[5] See generally Rochet and Tirole (2003), and Parker and Van Alstyne (2002).

TABLE 1 Sources of revenue in selected two-sided platforms

Industry	Two-Sided Platform	Category	Side One	Side Two	Side that "Gets Charged Least"	Sources of Revenue
Real Estate	Residential Property Brokerage	Market-makers	Buyer	Seller	Side One	Real estate brokers derive income principally from sales commissions.[a]
Real Estate	Apartment Brokerage	Market-makers	Renter	Owner/ Landlord	Typically Side One	Apartment consultants and locater services generally receive all of their revenue from the apartment lessors once they have successfully found tenants for the landlord.[b]
Media	Newspapers and Magazines	Audience makers	Renter	Advertiser	Side One	Approximately 80 percent of newspaper revenue comes from advertisers.[c]
Media	Network Television	Audience makers	Viewer	Advertiser	Side One	For example, the FOX television network earns its revenues primarily from advertisers.[d]
Media	Portals and Web Pages	Audience makers	Web "Surfer"	Advertiser	Side One	For example, Yahoo! earns 75 percent of its revenues from advertising.[e]
Software	Operating System	Demand coordinators	Application User	Application Developer	Side Two	For example, Microsoft earns at least 67 percent of its revenues from licensing packaged software to end-users.[f]
Software	Video Game Console	Demand coordinators	Game Player	Game Developer	Neither – Both sides are significant sources of platform revenue	Both game sales to end users and licensing to third party developers are significant sources of revenue for console manufacturers. Console manufacturers have sold their video game consoles near or below marginal cost (not taking into account research and development). Microsoft, for instance, is selling its Xbox for at least $125 below marginal cost.[g]
Payment Card System	Credit Card	Demand coordinators	Cardholder	Merchant	Side One	For example, in 2001, American Express earned 82 percent of its revenues from merchants, excluding finance charge revenue.[h]

Notes: [a]See Bureau of Labor Statistics, U.S. Department of Labor (2003) "Real Estate Brokers and Sales Agents," in *Occupational Outlook Handbook, 2002-03 Edition*, 362-364, http://www.bls.gov/oco/ocos120.htm. [b]See Ronan (1998). [c]See George and Waldfogel (2000). [d]See FOX (2000) "Annual Report". [e]See Yahoo! Inc. (2001) "Annual Report." [f]See IDC (1994) *1994 Worldwide Software Review and Forecast*, Report #9358, November, IDC (1995) *1995 Worldwide Software Review and Forecast*, Report # 10460, November, IDC (1996) *1996 Worldwide Software Review and Forecast,* Report #12408, November, IDC (1997) *1997 Worldwide Software Review and Forecast,* Report #14327, October, IDC (1999) *1999 Worldwide Software Review and Forecast,* Report #20161, October, IDC (2001) *Worldwide Software Market Forecast Summary, 2001-2005,* Report #25569, September. [g]See Becker (2000), Fahey (2002). [h]See American Express Company (2001) "Annual Report," http://www.onlineproxy.com/amex/2002/ar/pdf/axp_ar_2001.pdf.

The remainder of this chapter examines several multi-sided platform industries. It is helpful to divide multi-sided platforms into three categories: (1) market-makers; (2) audience-makers; and (3) demand coordinators. Table 1 provides further examples of multi-sided platform markets and businesses that participate in these markets. While by no means exhaustive, it illustrates the variety of multi-sided platform industries.

Market-makers enable members of distinct groups to transact with each other. Each member of a group values the service more highly if there are more members of the other group – because that increases the likelihood of a match and reduces the time it takes to find an acceptable match. Examples include exchanges such as NASDAQ and eBay, shopping malls such as those that dot the New Jersey Turnpike, and dating services such as Yahoo! Personals.[6]

Audience-makers match advertisers to audiences. Advertisers value the service more if there are more members of an audience who will react positively to their messages; audiences value the service more if there are more useful messages (Goettler, 1999). Advertising-supported media such as magazines, newspapers, free television, yellow pages, and many Internet portals are audience makers.[7]

Demand-coordinators make goods and services that generate indirect network effects across two or more groups. They are a residual category but economically the most interesting and the least studied. These platforms do not strictly sell "transactions" like a market maker or "messages" like an audience-maker. Software platforms such as *Windows* and the *Palm OS,* payment systems such as debit cards, and mobile telephones are examples (Rochet and Tirole, 2003).[8]

III. BUSINESS MODELS IN MULTI-SIDED PLATFORM MARKETS

Several issues occur repeatedly in multi-sided platform markets: getting both sides on board; balancing interests; multi homing; scaling and liquidity.

[6] See NASDAQ (2003) "Market Characteristics" http://www.nasdaq.com/about/market_characteristics.pdf, eBay Inc. (2003) "Company Overview," http://pages.ebay.com/community/aboutebay/overview/index.html, Pashigian and Gould (1998), and Yahoo! Inc. (2003) "Yahoo! Personals," http://personals.yahoo.com/.

[7] "In a fundamental sense, what advertisers demand, and what the various advertising media outlets supply, are units of audience for advertising messages. Thus advertiser demand for space in the print media and time in the broadcast media is a derived demand stemming from a demand for audience, and is a positive function of the size and quality of audience" (Ferguson, 1983).

[8] Mobile telephone services are not a one-sided market because most users both make and receive calls. A high termination charge raises the marginal cost of calls and lowers the marginal cost of call receptions. For a given call, end users are on a single side and their consumption behaviors depend on their own price (calling price for the caller and receiving price for the receiver). As a consequence, the choice of termination charge is not neutral. See Jeon, et al. (2001) for more detail.

A. Getting Both Sides on Board

An important characteristic of two-sided markets is that the demand on each side tends to vanish if there is no demand on the other – regardless of what the price is. There are many references in the literature on the firms discussed earlier about solving the chicken-and-egg problem (Gawer and Cusumano, 2002). For example, there would be no demand by households for payment cards if they could not use them anywhere and no demand by retailers for payment cards if no one had them. Which comes first – the cardholder or the retailer (Evans and Schmalensee, 1999)? Men will not go to dating clubs that women do not attend because they cannot get a date. Merchants will not take a payment card if no customer carries it because no transaction will materialize. Computer users will not use an operating system that does not have applications they need to run. Sellers of corporate bonds will not use a trading mechanism that does not have any buyers. In all these cases, the businesses that participate in these industries have to figure out ways to get both sides on board. Investment and pricing strategies are keys to getting both sides on board.

One way to do this is to obtain a critical mass of users on one side of the market by giving them the service for free or even paying them to take it. Especially at the entry phase of firms in multi-sided markets, it is not surprising to see precisely this strategy. Diners Club gave its charge card away to cardholders at first – there was no annual fee and users got the benefit of the float. Netscape gave away its browser to most users to get a critical mass on the computer user side of the market; after Microsoft started giving away its browser to all users, Netscape followed suit (Wong, 1998). Microsoft is reportedly subsidizing the sales of its Xbox hardware to consumers to get them on board (Becker, 2002). For monopoly and duopoly cases, if there is sufficient difference in the valuation of a transaction then the market with low valuation will be flooded in equilibrium and that type will pay zero price according to some recent theoretical work (Schiff, 2003).

Another way to solve the chicken-and-egg problem is to invest in one side of the market to lower the costs to consumers on that side of participating in the market. Microsoft provides a good example of this. As we saw earlier, it invests in applications writers by developing tools that help them write applications and providing other assistance that makes it easier for developers to write applications using Microsoft operating systems. To take another example, bond dealers take positions in their personal accounts for certain bonds they trade. They do this when the bond is thinly traded and the long time delays between buys and sells would hinder the market's pricing and/or liquidity. By investing in this manner, two-sided intermediaries are able to cultivate (or even initially supply) one side, or both sides, of their market in order to boost the overall success of the platform.

Providing low prices or transfers to one side of the market helps the platform solve the chicken-and-egg problem by encouraging the benefited group's participation – which in turn, due to network effects, encourages the non-benefited group's participation. Bernard Caillaud and Bruno Jullien (2001) refer to this strategy as "divide-and-conquer." Another effect of providing benefits to one side is that this assistance can discourage use of competing two-sided firms. For example, when Palm provides free tools and support to PDA applications software developers, it encourages those developers to write programs that work on the Palm OS platform, but it also induces those developers to spend less time writing programs for other operating systems (See discussion of Palm below.)

B. Pricing Strategies and Balancing Interests

Firms in mature multi-sided markets – i.e. those that have already gone through the entry phase in which the focus is on solving the chicken-and-egg problem – still have to devise and maintain an optimal pricing structure. In most observed multi-sided markets, companies seem to settle on pricing structures that are heavily skewed towards one side of the market in the sense that the margin (price less marginal cost as a percent of price) is far less on one side than the other. Table 1 summarizes the pricing structure for some multi-sided markets. For example, in 2001, excluding finance charge revenue American Express earned 82 percent of its revenues from merchants.[9] Microsoft earns the preponderance of its revenue from Windows from licensing Windows to computer manufacturers or end-users.[10] Real estate brokers (for sales as opposed to rent) usually earn most or all of their revenues from the sellers. Political tensions can also manifest themselves – looking out for their narrow interests, customers on each side of the market would like the other side to pay more. This is a familiar problem in the payment-card industry – in Europe a retailers association asked the European Commission to force the card associations to eliminate the interchange fee and this tension was behind the retailer litigation that was recently settled in the United States.[11]

Discerning the optimal pricing structure is one of the challenges of competing in a multisided market. Sometimes all the platforms converge on the same pricing strategy. Microsoft, Apple,

---

9 If finance charge revenues are included, American Express earned 62 percent of its revenues from merchants in 2001 (American Express Company (2001) "Annual Report," http://www.onlineproxy.com/amex/2002/ar/pdf/axp_ar_2001.pdf.).

10 From 1988 through 2000, Microsoft earned at least 67 percent of its revenues from licensing packaged software (such as Windows and Office) to end-users (either directly at retail or through manufacturer pre-installation on PCs). See IDC (1994) *1994 Worldwide Software Review and Forecast*, Report #9358, November, IDC (1995) *1995 Worldwide Software Review and Forecast*, Report # 10460, November, IDC (1996) *1996 Worldwide Software Review and Forecast*, Report #12408, November, IDC (1997) *1997 Worldwide Software Review and Forecast*, Report #14327, October, IDC (1999) *1999 Worldwide Software Review and Forecast*, Report #20161, October, IDC (2001) *Worldwide Software Market Forecast Summary, 2001-2005*, Report #25569, September. Note that the 67 percent figure underestimates the amount of revenue Microsoft earns from end-users because the other third of revenue coming from "Applications Development and Deployment" includes some end-user revenues as well. For example, database products used by business IT departments are included in the Applications Development category.

11 See British Retail Consortium (2002) "Retail Disappointment as Visa Exempted from EU Competition Rules," Press Release, July 25, http://www.brc.org.uk/Archive.asp, Guerrera (2002), MacKintosh (2001), and Pacelle, Sapsford and Scannell (2003).

IBM, Palm and other operating system companies could have charged higher fees to applications developers and lower fees to end-users. They all discovered that it made sense to charge developers relatively modest fees for developer kits and, especially in the case of Microsoft, to give a lot away for free. Nevertheless, Microsoft is known for putting far more effort into the developer side of the business than the other operating system companies (Gawer and Cusumano, 2002).

The debit card is an example in which different platforms made different pricing choices. In the late 1980s, the ATM networks had a base of cardholders who used their cards to withdraw cash or obtain other services at ATMs. They had no merchants that took these cards. To add debit services to existing ATM cards, the ATM networks charged a small interchange fee than the card associations charged (8 cents per transaction on a typical $30 transaction compared) to encourage merchants to install PIN pads that could read the ATM cards that cardholders already had and accept the pins they used to access the ATM machines (Evans and Schmalensee, 1999). Many merchants invested in the PIN pads – the number of PIN pads increased from 53,000 in 1990 to about 3.6 million in 2001.[12] The credit-card associations had a base of merchants who took their cards but it did not have cards that, like the ATM cards, accessed consumers' checking accounts. The credit-card systems imposed a much higher interchange fee than the ATM networks, about 37 cents versus 8 cents on a typical $30 transaction.[13] They did this to persuade banks to issue debit cards and cardholders to take these cards.[14] The number of Visa debit cards in circulation increased from 7.6 million in 1990 to about 117 million in 2001.[15]

Two other factors influence the pricing structure. There may be certain customers on one side of the market – Rochet and Tirole refer to them as "marquee buyers" – that are extremely valuable to customers on the other side of the market. The existence of marquee buyers tends to reduce the price to all buyers and increases it to sellers. A similar phenomenon occurs when certain customers are extremely loyal to the two-sided firm – perhaps because of long-term contracts or sunk-cost investments. For example, American Express has been able to charge a relatively high merchant discount as compared to other card brands, especially for their corporate card, because merchants viewed the American Express business clientele as extremely attractive. Corporate expense clients were "marquee" customers that allowed American Express to raise its prices to the other side of the market, merchants. In contrast, when the ATM systems entered into debit, they had "captive" cardholders – ATM cards could be used for debit transactions, so consumers did not need to be courted to accept the new payment form. Therefore, it has been the merchants – who must

[12] See Evans and Schmalensee (1999) and The Nilson Report (2002) No. 759, March.

[13] The ATM systems typically charged a flat interchange fee per transaction, while the interchange fee set by Visa and MasterCard varied with the size of the transaction. The reported interchange fee comparison is from 1998, around the time of substantial growth in debit for the ATM and credit-card systems (Evans and Schmalensee, 1999).

[14] Visa attracted consumers through an effective advertising campaign and attracted issuers through heavy investment in a debit processing facility, among other strategies (Evans and Schmalensee, 1999).

[15] See The Nilson Report (2002) No. 760, March and The Nilson Report (1991) No. 500, May.

purchase and install expensive machinery in order to process online debit transactions – who have been courted, as we saw above.

C. Types of Platform Market Structures

Several different multi-sided market organizations appear in practice. (1) Coincident platforms: several multi-sided platforms offer substitutable products or services on the same sides. That is the case in video games, operating systems, and payment cards. (2) Intersecting platforms: several n-sided platforms offer products or services that are substitutable on less than n sides. Browsers were sides of operating system and internet portal businesses. ATM networks do not support credit cards or other cards that are not linked to the depository institution while credit card systems do not really offer ATM cards. (3) Monopoly platforms that have no competition on any side. Although this could exist in theory, of course, it is hard to identify any industry for which this has been true (yellow pages was an example for a time perhaps in some places).

Another aspect of platform competition concerns the extent to which users rely contemporaneously on more than one platform for a side. Users may be dedicated to one platform because it is not efficient or otherwise beneficial to use more than one network. For example, most users do not want to use more than one operating system on their personal digital assistant. Users may also find that it is beneficial and efficient to use several competing platforms – that situation has been called multihoming. Most merchants accept payment cards from several competing card systems.

Platform industries often have multihoming on at least one side. Table 2 presents a summary. Consider, for example, personal computers. One could consider the two sides as consisting of personal computer end-users and developers of applications. The end-users do not multihome. They almost always use a single operating system and by far the preponderance of them use a Microsoft operating system.[16] The developers do multihome. According to Josh Lerner, in 2000, 68 percent of software firms developed software for Windows operating systems, 19 percent for Apple computers operating systems, 48 percent for Unix operating systems including Linux, and 36 percent and 34 percent for proprietary non-Unix operating systems that run on minicomputers and proprietary operating systems that run on mainframes respectively.[17] In fact, in recent years the percentage of software firms developing for non-Microsoft operating systems has increased. The fastest-growing category has been software firms developing for Unix operating systems including Linux. The percentage of developers in this category increased from 29 percent in 1998 to 48 percent in 2000.[18]

[16] IDC (2001) *Worldwide Software Market Forecast Summary, 2001-2005,* Report #25569, September.
[17] The percentages total to 205, indicating substantial multihoming on the part of developers. See Lerner (2002) and Corporate Technology Directory (1990-2000).
[18] See Lerner (2002) and Corporate Technology Directory (1990-2000).

TABLE 2 The presence of multihoming in selected two-sided platforms

Two-Sided Platform	Side One	Presence of Multihoming for Side One	Side Two	Presence of Multihoming for Side Two
Residential Property Brokerage	Buyer	*Uncommon:* Multihoming may be unnecessary, since an MLS allows buyers to see property listed by all member agencies.[a]	Seller	*Uncommon:* Multihoming may be unnecessary, since an MLS allows the listed property to be seen by all member agencies' customers.[a]
Securities Brokerage	Buyer	*Common:* The average securities brokerage client has accounts at three firms.[b] Note that clients can be either or both buyers or sellers.	Seller	*Common:* The average securities brokerage client has accounts at three firms.[b] As mentioned, clients can be either or both buyers or sellers.
Business-2-Business	Buyer	*Varies:* For example, multihoming may be unnecessary for some online B2B sites, since buyers can go directly to the B2B platform instead of contacting multiple individual suppliers.[c]	Seller	*Varies:* Multihoming may be unnecessary since the B2B can inexpensively reach a large audience.[d]
Peer-2-Peer	Buyer	*Varies:* Multihoming may be unnecessary for buyers using online auction sites since eBay holds 85% of the market share (i.e. it seems that most people purchase their online auction products at eBay). Alternatively, multihoming may be more common for online dating services where there are many sites and a large audience of online singles (considered to be available singles, as opposed to buyers).[e]	Seller	*Varies* – Multihoming may be unnecessary for sellers using online auction sites since eBay holds 85% of the market share (i.e. it seems that most people auction their products at eBay). Alternatively, multihoming may be more common for online dating services where there are many sites and a large audience of online singles (considered to be available singles, as opposed to sellers).[e]
Newspapers and Magazines	Reader	*Common:* In 1996, the average number of magazines issues read per person per month was 12.3.[f]	Advertiser	*Common:* For example, Sprint advertised in the New York Times, Wall Street Journal, and Chicago Tribune, among many other newspapers, on Aug. 20, 2002.[g]
Network Television	Viewer	*Common:* For example, Boston, Chicago, Los Angeles, and Houston, among other major metropolitan areas, have access to at least four main network television channels: ABC, CBS, FOX, and NBC.[h]	Advertiser	*Common:* For example, Sprint places television advertisements on ABC, CBS, FOX, and NBC.[i]
Operating System	Application User	*Uncommon:* Individuals typically use only one operating system.[j]	Application Developer	*Common:* As noted earlier, the number of developers that develop for various operating systems indicates that developers engage significant multihoming.[k]
Video Game Console	Game Player	*Varies:* The average household (that owns at least one console) owns 1.4 consoles.[l]	Game Developer	*Common:* For example, Electronic Arts, a game developer, develops for Nintendo's GameCube, Microsoft's Xbox, and Sony's Playstation 2, among other consoles.[m]
Payment Card	Cardholder	*Common:* Most American Express cardholders also carry at least one Visa or MasterCard.[n]	Merchant	*Common:* American Express cardholders can use Visa and MasterCard at almost all places that take American Express.[n]

Notes: [a]See Frew and Jud (1987); [b]See Scherreik (2002); [c]See Lucking-Reilly and Spulber (2001); [d]See Federal Trade Commission Staff (2000) "Entering the 21st Century: Competition Policy in the World of B2B Electronic Marketplaces: Efficiencies of B2B Electronic Marketplaces," October, http://www.ftc.gov/os/2000/10/index.htm#26; [e]See Cisneros (2000) and Festa (1996); [f]See FCB (1998) "Magazines in the Information Age," *Media Research Report,* Spring, http://www.magazine.org/resources/research/fcb_magazines_infoage.html; [g]*See New York Times* (2002), *Wall Street Journal* (2002), and *Chicago Tribune* (2002); [h]See ABC (2002) "Local Stations," http://abc.abcnews.go.com/site/localstations.html., CBS (2002) "CBS Info," http://www.cbs.com/info/hdtv/., FOX (2002) "Fox Affiliates," http://www.fox.com/links/affiliates.htm., and NBC (2002) "Local Stations," http://www.nbc.com/nbc/header/Local_Stations/; [i]See Fisher (1997); [j]See Hacker (2001); [k]See Lerner (2002) and Corporate Technology Directory (1990-2000); [l]See Yankee Group (2001) "Video-Game Penetration Grows to 36 Million Households in 2001," Reuters News, November 19, http://about.reuters.com/newsreleases/art_19-11-2001_id785.asp; [m]See "Game Makers Hedge Bets in Console Wars" (2001) *USA Today*, November 16, http://www.usatoday.com/life/cyber/tech/review/games/2001/11/19/game-makers.htm; [n]See Evans and Schmalensee (1999).

Multihoming and intersecting platforms affect both the price level and the pricing structure. Theory and empirics are not far enough advanced to say much more.

D. Scaling and Liquidity

Successful multi-sided firms such as Microsoft, eBay, Yahoo!, and Diners Club, have taken time to test and tweak their platforms to build liquidity before making major investments. These firms established presence in small markets first, and used trial and error to identify the *correct* technology and operations infrastructure in which to invest. Many successful multisided firms seem to adopt a fairly gradual entry strategy in which they scale up their platform over time. Though much of the network economics literature may suggest that the multisided firm should rely on the right initial investments to help build liquidity over time, it is often difficult to predict just what the right technology and operations infrastructure will be. Therefore, successful multi-sided firm seem to found it advantageous to establish efficient buy-seller transactions first, and make large investments only after the platform has been tested. For instance, eBay expanded outside the collectibles market only when users started listing such items up for sale. Figure 1 below outlines the growth of eBay's sales categories. It shows the gross merchandise sales per category in the third quarter of 2002 broken down by category. The three categories – collectibles, early practicals, and practicals – are grouped by the year they debuted.

FIGURE 1 eBay's category growth[19]

Category	Q3-02 *(millions)*	Yr/Yr	
Collectibles:			
Collectibles	$920	11%	Debuted First ~ 1995
Coins & Stamps	$400	51%	
Early Practicals:			
Computers	$1,600	44%	Debuted Second ~1997
Consumer Electronics	$1,400	78%	
Books, Movies, Music	$1,200	44%	
Sports	$1,100	35%	
Practicals:			
Motors	$3,800	116%	Debuted Third ~ 1999
Clothing	$700	98%	
Home & Garden	$470	90%	

[19] See eBay Inc. (2001) "Annual Report."

Palm also did not invest large amounts into developer support until it achieved sufficient liquidity on the user side of the market. While small investments were made in releasing a software development kit (SDK) with the Palm Pilot and making the Palm OS and architecture accessible to outside developers, Palm made few efforts to sign up key development partners or to support the Palm development community through classes, conferences, and other community support activities until a critical mass of users was assembled in 1998. (See discussion of Palm below.)

Many successful multi-sided firms have tested and modified their platforms with minimal investment and then scaled up according to what works best. An example comes from Yahoo!, a firm that has, over time, experimented with providing pages geared to different audiences such as Yahoo!Personals, Yahoo!Finance, and Yahoo!Travel. In 1996, Yahoo! developed Yahooligans!, a directory site that linked to content that appealed to children.[20] Yahoo! invested only one programmer and one business developer in their early efforts on Yahooligans!, both working half time for three months before the launch. As it turned out, the bandwidth available to most children at the time – 28 kbps – was too slow to allow for content that kept kids' attention. Yahooligans! was a relative failure among Yahoo! pages at the time, but that failure came at a minimal loss to Yahoo! (Bergin, 1998).

A final observation on business models. Contrary to the traditional economic theory of network effects (Arthur, 1989) and the business advice based on that theory (Shapiro and Varian, 1999), 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 in these industries ultimately did not retain the leadership position: Diners Club in cards, Apple in personal computer, Apple in hand-held devices, and OnSale in on-line exchanges. Also, as noted above, despite network effects many platform industries have several overlapping competing platforms and most have multi-homing on at least one side.

IV. DINERS CLUB AND PRICING IN THE PAYMENT CARD INDUSTRY

Frank McNamara, the president of a New York credit company, was having lunch in Manhattan in 1949. A year later he had a thriving business based on this experience. According to *Newsweek*:[21]

> "Halfway through his coffee, McNamara made a familiar, embarrassing discovery; he had left his wallet at home. By the time his wife arrived and the tab had been settled, McNamara was deep in thought." Result: the "Diners Club," one of the fastest-growing service organizations.

[20] See http://www.yahooligans.com
[21] "Dining on the Cuff" (1951) Newsweek, January 29.

McNamara and an associate, Ralph Schneider, started small. Beginning with less than $1.6 million (in 2002 dollars)[22] of start-up capital they signed up 14 New York restaurants, charging them 7 percent of the tab, and initially gave cards away to people.[23] (The observant reader will note that for most of us owning a credit card would not help our predicament when we leave our wallets at home. The proverbial apple falling on the head is a mysterious force.)

By its first anniversary there were 42,000 cardholders who were each paying $18 year for membership in the "club." There were 330 U.S. restaurants, hotels and nightclubs that accepted these cards; they paid an average of 7 percent of the cardholder's bill to Diners Club. In March 1951, Diners Club handled about $3 million of exchanges between cardholders and merchants (it reportedly made almost $60,000 in profit before taxes).[24] In 1951 Frank McNamara, founder of Diners Club, predicted that monthly transaction volume would increase to about $7 million by the end of the year.[25] Diners Club was therefore getting more than three quarters of its revenues from merchants.[26] Moreover, the margin on cardholders was low since they were getting free float for an average of two weeks.

By 1958, it had raised the cardholder fee to about $26 but left the merchant fee at 7 percent. Nevertheless, it continued to earn most of its revenues, about 70 percent, from merchants.[27] The same year saw the birth of two of the major competitors to Diners Club. American Express, which had long been in the travel and entertainment business with its travelers checks and travel offices, decided to enter the industry. It recognized need to get both sides on board. Even before the first transaction, it had already acquired a cardholder base of 150,000 from the American Hotel Association's card program, as well as the program's 4,500 hotels for its merchant base. It bought another 40,000 cardholders from *Gourmet* magazine's dining card program (Grossman, 1987). It soon had a cardholder base to sell to T&E merchants, and sent out representatives to sign them up. It also advertised in newspapers for additional cardholders (Grossman, 1987). American Express's positive reputation in the T&E industry was also a benefit in attracting both sides of the business (Grossman, 1987). By the time of its entry on October 1958, it already had 17,500 merchant locations and 250,000 cardholders (Grossman, 1987).

American Express adopted a slightly different pricing policy than Diners Club. It initially set its annual fee at $31, $5 higher than Diners Club, thereby suggesting that it was the more "exclusive" card (Friedman and Meehan, 1992). But it set the initial merchant discount slightly lower: 5 to 7 percent for restaurants; and 3 to 5 percent for the recalcitrant hotel industry.[28] Within a year, its

---

[22] Except where noted, all dollar figures in this paper have been adjusted to constant 2002 dollars.
[23] See Mandell (1990) and Grossman (1987).
[24] "Charge It, Please" (1951) Time, April 9.
[25] "Dining on the Cuff" (1951) Newsweek, January 29.
[26] Cardholder fees totaled about $745,000 while merchant fees totaled about $2.5 million.
[27] "On-the-Cuff Travel Speeds Up" (1958) Business Week, August 16.
[28] "On-the-Cuff Travel Speeds Up" (1958) Business Week, August 16.

cardholder base had grown to 700,000 and its merchant base to 37,000.[29] With its slightly higher cardholder fee and slightly lower merchant fee, American Express received just under 55 percent of its revenue from merchants, compared to over 65 percent for Diners Club in 1959.[30] This would grow over time, however, as spending per card increased faster than annual fees.

American Express had been successful getting both sides on board but initially struggled to make a profit. It responded by putting more pressure on cardholders to pay promptly. It also managed to raise cardholder fees without suffering significant attrition. There were 900,000 cardholders that could use their cards at 82,000 merchant locations by the end of 1962, the first year that the card operation posted a small profit (Hammer, 1962).

The other significant entrant in 1958 was Bank of America (this card evolved into Visa). Unlike many other banks, it had always focused on lending money to the middle class. By 1958, Bank of America had extensive experience in making small loans on consumer durables such as refrigerators and automobiles. It had also become the largest bank in the country (Mandell, 1990).

One of its small competitors, the First National Bank of San Jose, started a charge card in 1953. At the time, Bank of America considered introducing its own card but decided that there was not a good enough business case. After studying the emerging industry over the next few years they decided to introduce a credit card. Credit-worthy customers would receive cards with limits of either $1,600 or $2,600; prior authorization would be required for purchases over $130; and a revolving credit option was available for some cardholders (Wolters, 2000). Revolving credit was the feature that distinguished this card from existing charge cards. The merchant fee was initially set at 5 percent and lowered soon thereafter (Wolters, 2000). Cardholders did not pay an annual fee but paid finance charges on any revolved balances at an annual interest rate of 18 percent (Wolters, 2000). This was a more merchant-friendly balance than that set by Diners Club or American Express, especially considering that Bank of America's card was also extending credit on a revolving basis.

Bank of America did a market test in Fresno in the fall of 1958. Three hundred retailers signed up initially, and every Bank of America customer in the Fresno area received a card. According to one study, "This mass mailing of 60,000 cards had been William's (the Bank of America leader of the effort) solution to the problem of how to convince retailers that enough individuals would possess a card to make their participation in the program worthwhile. His solution worked, for during the next five months another eight hundred Fresno-area retailers joined the newly named 'BankAmericard' program" (Wolters, 2000). They expanded throughout the state during the following year. By the end of 1959, 25,000 merchants accepted the card and almost two million California households had one.

Things did not go well. Fraud was rampant. The number of delinquent accounts was five times higher than expected. Large retailers resisted joining. The program lost $45 million in 1960. The bank worked on collection problems and lowered the merchant fee to as low as three percent to

[29] "Towards an Ever-Fuller Life on Credit Cards" (1959) Newsweek, September 28.
[30] Figures taken from Newsweek 1959 article, and assuming average of 5 percent for Amex per Newsweek 1958 article.

entice retailers. Delinquencies declined and the merchant base increased to 35,000 in 1962. It was profitable by the early to mid 1960s.

One notable common theme across the experiences of the three startups is the importance of getting both sides on board. Diners Club, starting from scratch and trying to sell a new product, had to build slowly. It set a relatively cardholder-friendly balance, collecting most of its revenues from merchants. This made it possible to sign up cardholders even before a significant merchant base had been signed. The cardholder base could then be used, of course, to sign up merchants. By the time of American Express's entry, the idea of a charge card had already been established by Diners Club and others. American Express was able to capitalize on that by entering at a larger scale. Indeed, it probably did not have the option of entering as gradually as Diners Club did, since what cardholders or merchants would sign up with a fledgling American Express card given the option of an established alternative. American Express therefore bought up customer bases on both sides before it entered. It was also able to use its established position in the T&E business to set a slightly higher cardholder fee, to cultivate the upscale image that served it well for a long time.

Bank of America also developed a large cardholder base in entering the business. It used its base of depository customers, the largest in California, to develop an instant cardholder base, which it then used to sign up merchants. Even with its prominent position among consumers in California, however, Bank of America had to strike a more merchant-friendly balance, lowering its merchant fees until it could get enough merchants on board.

Perhaps as a consequence of entering at a larger scale, both American Express and Bank of America, unlike Diners Club, had a hard time making a profit early on. Diners Club made a profit in its first year, while the other two firms did not see a profit until their fourth or fifth years. It took time to weed out delinquent cardholders. Having to make an upfront investment in building a profitable cardholder base is a lesson that holds even for new card issuers today (although they can join Visa or MasterCard and do not have to build an entire system). There were also the operational problems that would be expected in starting up any new business. Bank of America had such (misplaced) confidence its customers would pay their bills that it did not even set up a collections department (Nocera, 1994).

V. THE PALM OS AND COMMUNITY-BUILDING IN SOFTWARE PLATFORMS

Palm was founded in 1992 as an effort to develop software applications for personal digital assistants (PDAs). In 1993, Palm released the Zoomer in collaboration with five other companies – Casio, GeoWorks, Tandy, Intuit, and America Online. Palm was responsible for the application development while the other five companies handled the hardware, the operating system, and distribution. The Zoomer, with its high price ($820) and poor handwriting recognition, ultimately failed with only 15,000 units sold (Gawer and Cusumano, 2002). Zoomer was not the first failure.

Apple had made the first attempt with its Apple Newton in 1993. Other entrants into the PDA market from 1993 to 1996 were the Psion Series 3, Hewlett-Packard LX, Hewlett-Packard OmniGo, Sharp Zaurus, Sharp Wizard, and Microsoft's Windows CE[31] PDAs. With the exception of the Windows CE PDA, all followed an integrated strategy where the PDA manufacturer produced the hardware, operating system, and applications. Windows CE PDAs had a non-integrated strategy, where the hardware, operating system, applications were all developed by different firms (although Microsoft did end up writing the operating system as well as many applications for Windows CE). The integration of application and operating system development within one firm was common in the software industry, for example Microsoft produced some of the most popular applications for the Windows OS (Evans, 2003).

Palm re-entered the handheld market in 1996 with the Palm Pilot. This integrated hardware, an operating system, and some applications. Palm overcame the handwriting recognition hurdle with the invention the Graffiti text entry method, providing users a simple method to enter text. Industry experts described Graffiti as the "killer app" for the Palm Pilot (Feldstein and Flanagan, 2001). The pricing of the Palm Pilot also reflected Palm's desire to have deep market penetration. While 3Com, Palm's parent company from June 1997 to March 2000, wanted Palm to price the Palm Pilot high to extract profits, Palm insisted that prices be kept low to expand the user base (Gawer and Cusumano, 2002). The Palm Pilot's combination of low price and robust features were a hit with consumers and provided the "most bang for the buck" (Atluru and Wasserstein, 1998). The Palm Pilot met immediate success with over 360,000 units sold in 1996, representing 51 percent share of the market (Atluru and Wasserstein, 1998).[32]

Palm has gone through three stages in developing what is now a three-sided platform business. In the first phase from about 1996 to 1998, it entered as an integrated platform. It made the hardware, operating system, and applications for the Palm Pilot and integrated them together. It planned to court developers, however, only after it had a significant user base. According to Donna Dubinsky, CEO and one of its founders, "We are a highly integrated product that delivers end user results. We are not having a developer conference until we sell a million units" (Atluru and Wasserstein, 1998). A developer conference is one of the major methods used by software platforms to stimulate the writing of applications. Nevertheless, even during this initial phase Palm laid the groundwork for getting a developer community on board.[33] In early 1996, Palm released its first software development kit (SDK), which included the source code for the Palm Pilot's bundled applications. This source code served as a model that outside developers could reference to build other applications (Gawer and Cusumano, 2002).

[31] Windows CE was renamed Pocket PC at version 6.0.

[32] This represents market share of Palm hardware PDA and does not include licensed Palm OS devices.

[33] In fact, in the first year of release, the Palm attracted over 2,000 third-party developers to its platform without any developer support efforts (Atluru and Wasserstein, 1998).

In 1998, Palm sold over 2 million PDAs and launched the second phase of market development. Palm invested resources in persuading developers to write applications for the Palm OS. It offered business development resources to developers, including joint development, marketing, and bundling. Palm's most important business development resource was its software development forums (SDF), which helped external developers start and grow their own businesses. The SDF offered advisory meetings, business seminars, workshops, and networking events. It also started a $54 million venture capital unit called Palm Ventures to support businesses focusing on Palm OS applications. The company also offered Palm OS development classes regularly, and encouraged other activities among its community of users through developer portals (Gawer and Cusumano, 2002). The strategy to actively get developers on board was successful. The number of registered developers grew from 7,500 in 1998 to 220,000 in 2002.[34] Palm continued to secure its market leadership, with over 7 million units shipped by 2000.

The third phase, from about 2001 to the present, has moved Palm closer to being a pure operating system company. Palm has created alliances with Sony, Nokia, Handspring, IBM, Qualcomm, Supra, Symbol Technologies, Fossil[35], and TRG Products to create PDAs and other computer appliances based on the Palm OS (Gawer and Cusumano, 2002). Palm has reported that this greatly increases the market share of Palm OS based PDAs over their potential market share with an integrated strategy.[36] In addition, Palm OS licensing has expanded the scope of the Palm OS beyond traditional PDAs to such devices as wrist watches (Fossil Wrist PDA), cellular phones (Handspring Treo), and multimedia players (Sony Cliè). While Palm manufactured PDAs claimed 50 percent of the market in 2000, devices powered by the Palm OS operating system had over 75 percent share of the worldwide personal companion device market according to a report by International Data Corporation.[37] This market share increased to over 82 percent by the second half of 2002.[38] In January 2002 as Palm powered devices hit 20 million sold, Palm formed a Palm OS platform subsidiary, PalmSource, further separating the Palm OS from Palm's hardware solutions.[39]

Palm claims that its operating system was designed from the start to provide a platform for these three sides. According to Palm's web site, "From the start, the flexible, extensible Palm OS has been designed to grow and evolve in response to user needs. The open, modular architecture allows

---

[34] Palm, Inc. (1998) "Key Strategic Relationships And Growth To 7,500 Developers Propel 3Com Palm Computing® Platform Into Corporate America," June 15, http://www.palm.com/pr/enterprise.html; Palm, Inc. (2002) "Annual Report."

[35] Palm, Inc. (2002) "Fossil Signs License Agreement with Palm Source For First Ever Palm Powered Wrist Devices," November 18, http://www.palmsource.com/press/2002/111802.html.

[36] Palm, Inc. (2001) "Annual Report."

[38] Palm, Inc. (2003) "Palm Powered Share Increases in Second Half of 2002," January 27, http://www.palmsource.com/press/2003/012703.html.

[39] Palm, Inc. (2002) "Palm Completes Formation of Palm OS Subsidiary as Palm Powered Devices Hit 20 Million Sold," January 21, http://www.palmsource.com/press/2002/012102.html.

our more than 50,000 developers, licensees, alliance, and OEM partners to develop innovative new products and applications for a rapidly expanding global market" (Gawer and Cusumano, 2002).

Palm has subsidized both the application developer side of the market and the user side of the market by licensing their operating system at a loss and giving their development tools away for free. It is able to do this because of its profits in hardware sales. Palm provides an SDK, Palm OS Emulator, Palm OS Simulator, and sample source code free for download on its web site. Palm also provides links to third-party integrated development environments, both free and not, for download on their web site.[40] Palm also offers an online developer support program to offer a full range of development services. The Basic Level Membership is offered at no cost. The Advanced Level Membership is available for just $500 per year and includes direct technical, marketing and training support.[41] Only a small percentage of Palm's revenues were from licensing of the Palm OS. PalmSource revenues totaled $67, $88, and $50 million, while operating loss totaled $17, $8, and $8 million in fiscal years 2002, 2001 and 2000, respectively.[42] This continued loss of PalmSource represents Palm's continued efforts in trying to expand the user base of the Palm platform.

VI. SUMMARY

Multi-sided platform markets are becoming an increasingly important part of the economy. They range from relatively small emerging companies like eBay, Yahoo!, and Palm, to relatively large and mature companies like American Express. These markets also have had a large impact on the recent information technology boom, and undoubtedly, they will continue to be important, as internet based commerce expands its scope to include both new and old economy firms. In addition to internet commerce, other increasingly important industries, such as credit cards, operating systems, shopping centers, and mass media, are all governed by the economics of multi-sided platforms.

While it is not my intention to generalize from a small sample, and many issues need to be investigated in more depth, a number of business models appear common across both the cases studied here. First, differential pricing is used to get both sides on board. Second, once both sides are on board, pricing continues to play a key role in maintaining both sides of the platform. Third, multi homing often occurs in multi-sided platforms. And, fourth, one solution to the pricing complexity that has been used successfully is to start with a small but scalable platform.

---

[40] Palm, Inc. (2003) "PalmOS: Desktop Development," http://www.palmos.com/dev/tools.

[41] Palm, Inc. (2003) "PalmOS.com: Palm OS Developer Program," http://www.palmos.com/dev/programs/pdp.

[42] Palm, Inc. (2002) "Annual Report."

CHAPTER 3

How Catalysts Ignite: The Economics of Platform-Based Start-Ups

David S. Evans

ABSTRACT

Entrepreneurs who start multi-sided platforms must secure enough customers on both sides, and in the right proportions, to provide enough value to either group of customers and to achieve sustainable growth. In particular, these entrepreneurs must secure "critical mass" to ignite the growth of their platforms; the failure to achieve "critical mass" quickly results in the implosion of the platform. There are a number of strategies available to entrepreneurs to reach critical mass. For example, the "zigzag" strategy involves successive accretions of customers on both sides to build up the value to both. The relevant strategies depend in large part on whether the nature of the platform requires securing participation by both platform sides at launch (e.g. dating venues), whether it is possible to acquire one side before approaching the other side (e.g. search engines), and whether it is necessary to make precommitments to one side to induce them to make investments (e.g. video games).

I. INTRODUCTION

Starting a business and getting it to the point where it is economically viable is the most difficult problem for all entrepreneurs. Most new businesses fail. In the United States, 61 percent of new businesses that were started in second quarter of 1998 had ceased business within five years.[1] Venture capital firms that invested in new firms do not get any of their money back in 43.7 percent of the first-round investments they make and get less than their initial investment back in 66.7 percent of the first-round investments.[2]

The start-up problem is particularly difficult for firms that are based on multi-sided platforms. In addition to the usual problems faced by new firms they often must contend with the well-known chicken-and-egg problem. Their firm can deliver value to one side of the platform only if there are

[1] Amy Knaup and Merissa C. Piazza, "Business employment dynamics data: survival and Longevity, II," *Monthly Labor Review Online 130, no.9 (September 2007), http://www.bls.gov/opub/mlr/2007/09/art1full.pdf.*

[2] Andrew Metrik, *Venture Capital and the Finance of Innovation* (New York: Wiley, 2006), Exhibit 7-4.

participants on the other side of the platform. They have to figure out how to get both sides on board their platform. That problem, which is the subject of this chapter, is very different from that faced by a one-sided startup whose main challenge is getting just one set of customers to buy its product or service.[3]

The chicken-and-egg problem is central to the study of multi-sided platforms. Yet most of the theoretical and empirical research on two-sided businesses has focused on mature platforms and examined their pricing structures and other properties. Little attention has been given to critical issues that entrepreneurs must solve to create a viable platform business.[4] These include strategies for getting both sides on board, the role of critical mass in establishing the foundations for success, and the particularly thorny issues that arise when both sides must arrive simultaneously.

This introduction provides an overview of multi-sided platforms and the start-up problem. Section II then presents some building blocks for solving this problem. A few of these building blocks are similar to those for one-sided businesses while others are unique to multi-sided ones. Section III describes several complementary strategies for solving the chicken-and egg problem. We then turn to two case studies. Section IV examines a modern classic in a failed strategy for starting up multi-sided platforms: the en masse demise of the B2B exchanges whose assured success was extolled by many academics and practitioners.[5] Section V considers how social networking sites have started up and considers a successful and unsuccessful one. Section VI presents some concluding remarks.

A. The Catalyst Framework

We use the framework developed by Evans and Schmalensee.[6] A business is an "economic catalyst" if it creates value by bringing two or more groups of customers together and getting them to interact. Catalysts create value by reducing transactions costs faced by multiple distinct economic agents that would benefit from coming together. Catalysts reduce search efforts, fa-

---

[3] These comments also pertain to communication platforms in which it is important to distinguish between senders and receivers—even though they are the same people—for both economic analysis and business strategy.

[4] There are several notable exceptions to this statement. The role of critical mass in launching exchanges is analyzed in a dynamic model in Gabor Fath and Miklos Sarvaryi, "Adoption Dynamics in Buyer-Side Exchanges," *Quantitative Marketing and Economics*, Vol. 1, No. 3 (September), pp. 305-335. Also see Bernard Caillaud, and Bruno Jullien, (2003), "Chicken & Egg: Competition Among Intermediation Service Providers," *Rand Journal of Economics*, Vol. 34, No. 2 (Summer), pp. 309-328 which considers platform launch but in a single-period setting. The role of commitment is analyzed in a two-period game in Andrei Hagiu, "Pricing and Commitment by Two-Sided Platforms", *The RAND Journal of Economics* 37, no. 3 (fall 2006). A managerial treatment of the issues is presented in Annabelle Gawer and Michael Cusumano : How Companies Become Platform Leaders (Sloan Management Review, Winter 2008, Vol. 49, No. 2, pp.28-35.

[5] *See* David Lucking-Reiley and Daniel F. Spulber, "Business-to-Business Electronic Commerce," *Journal of Economic Perspectives 15, no.1 (2001):55-68*; Nicholas Carr, *The Digital Enterprise: How to Reshape Your Business for a Connected World*, (Massachusetts: Harvard Business Press, 2001); and Arthur Sculley and William Wood, B2B Exchanges: The Killer Application in the Business-to-Business Internet Revolution, (ISI Publications, 1999).

[6] *See* David S. Evans & Richard Schmalensee, *Catalyst Code: The Strategies Behind the World's Most Dynamic Companies* (Massachusetts: Harvard Business School Press, 2007).

cilitate matching, and make it easier for the two groups of economic agents to exchange value between each other. In the traditional literature, a catalyst is referred to as a "two-sided market" or as a "multi-sided platform."[7] In this chapter we use catalyst and multi-sided platform interchangeably.

The economic value created by the catalytic reaction is essential for understanding the feasible set of business strategies that a multi-sided platform can use. That value must be significant enough to warrant the cost and risk of investment in developing the platform. The value also provides the "pie" that can be split among the distinct groups of economic agents to provide incentives for them to join and interact on the platform with a portion of the pie going to the platform for performing its role. Some of the pie can be used to subsidize certain groups of economic agents, or members of those groups, to join the platform.

Catalyst innovators are ones who discover that it is possible to create economic value by getting two or more groups of economic agents together on a shared platform or develop a more efficient platform for starting and accelerating a catalytic reaction.

B. Securing Ignition

In chemical catalysis it is necessary to get the catalytic agent and the chemical agents in the right proportions to ignite and accelerate a reaction. The same is true for economic catalysis. Both economic agents have to be present on the platform in the right parts and levels to create any value at all and to accelerate value creation.

This conundrum is often referred to as the chicken-and-egg problem in the academic and popular literature on platforms. That analogy does not work for many platforms. The problem that platforms face is sometimes sequential as the riddle suggests—does the platform need to get economic agents ***A*** on board the platform before economic agents ***B***; or ***B*** before ***A***? Other times, though, it is simultaneous— how does the platform secure the participation of economic agents ***A*** and ***B*** so that both will be present on the platform when members from each group show up. That typically involves solving a very difficult coordination problem between the platform and these two groups of economic agents.

The problem, however, is not just getting members of the two groups of economic agents to show up at the same time to create value. There have to be enough members of group ***A*** to make it valuable to members of group ***B*** to incur the costs of participating in the platform and to return to it in subsequent periods; and vice versa. Strength in numbers arises primarily because economic agents in one group are searching for appropriate value-creating matches among members of the other group. There have to be enough members to make it likely that economic agents will find valuable matches.

[7] I prefer not to use the "two-sided market" term because the term really refers to a platform business that provide a matchmaking service in competition with other platform businesses in a market for that service.

In the literature on market microstructure, for example, for financial exchanges a "liquid"market is one that has enough buy and sell-orders to facilitate transactions.[8] Markets that are too "thin", or too illiquid, collapse. There is a critical mass of buy and sell orders that allows markets to sustain themselves. We will see below that the B2Bs failed because they did not achieve a critical mass of buyers and seller; in particular they did not create enough value to suppliers to entice them to participate in the exchanges.

C. Catalytic Ignition and Critical Mass

Figure 1 shows the basic concept of critical mass and catalytic ignition.[9] There is a range of minimal numbers of customers in each group that, if achieved, provides a "thick enough market" or a sufficiently "liquid" market to permit sustainable growth. When the mass of economic agents on either side is insufficient a catalyst fizzles rather than ignites. Once a catalyst achieves critical mass on C'-C", for example, it can grow to its profit-maximizing potential of D*; if it does not achieve critical mass on the segment C'-C" it contracts and fails.

The growth paths to critical mass depend on many factors including pricing. But the point here is that achieving critical mass is essential. Google Video, for example, failed to achieve critical mass because it did not generate enough content to attract viewers and did not attract enough viewers to attract paid or user-generated content.[10]

The optimal growth path to critical mass and to long-run equilibrium is well away from the horizontal and vertical axes in most plausible cases.[11] Relatively balanced growth is necessary. This is reflected in Figure 1 in that the equilibrium growth path to critical mass must occur within the triangle 0-C'-C". Having too many of one side and too few of another side will lead to quick failure.

The challenge that catalyst entrepreneurs face is how to achieve the critical mass that is necessary for ignition. That means getting to critical mass over some reasonable space of time. One can think of this phase as the ignition phase of the product launch process, in which customers are trying the platform and assessing its value; these early adopters will stop coming back, and stop recommending it to their

---

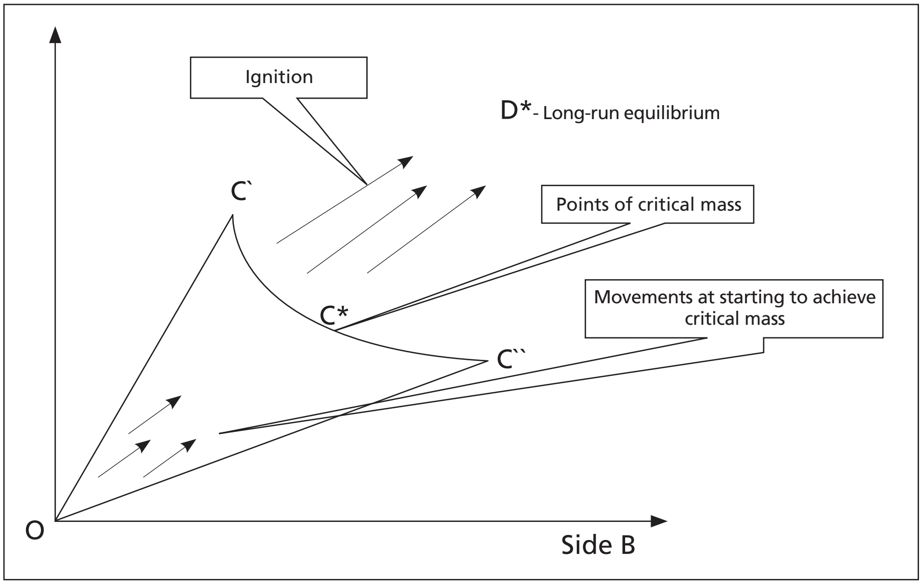
[8] In financial exchanges liquidity providers (such as market makers) provide buy and sell orders. They attract liquidity takers (such as investors who buy and sell through broker-dealers). In other exchange markets there is no institution that plays the role of liquidity providers. However, in these other exchange markets it is essential that the platform attract enough buyers and sellers to make the market liquid. *See* Larry Harris, *Trading and Exchanges: Market Microstructure for Practitioners* (Oxford University Press, 2002).

[9] For the technical development of this framework see David S. Evans and Richard Schmalensee, "Failure to Launch: Critical Mass in Platform Businesses,". We show that in certain plausible cases there is a critical level of liquidity that platforms need to get to in order to evolve towards a stable equilibrium with largest platform size. This framework does not cover the process of getting to critical mass. This chapter uses the framework to motivate an informal analysis of how catalysts can achieve critical liquidity.

[10] See Steve Johnson, "Google Video isn't ready for Prime Time," *Chicago Tribune,* January 18 , 2006, and David Pogue, "Google Video: Trash Mixed with Treasure," *The New York Times,* January 19th , 2006, http://www.nytimes.com/2006/01/19/technology/circuits/19pogue.html

[11] See Evans and Schmalensee, "Failure to Launch," op. cit.

FIGURE 1 Catalytic Ignition and Critical Mass



friends, if the platform does not grow quickly enough. The entrepreneur must increase the number of customers on each side to a point where there are enough to reach critical mass. That involves horizontal and vertical movements within the cone-shaped area in Figure 1. A transaction platform, for example, needs to take actions that increase buyers and sellers.[12] The entrepreneur must also maintain the right proportions of customers on each side so that there are enough participants on each side to interest the participants on the other side. That involves diagonal movements to the northeast in Figure 1. An exchange platforms needs to make sure there are enough buyers to interest sellers and vice versa.

II. BASIC CONCEPTS

There is an extensive literature on launching new products.[13] Key tactics involve logistics; advertising, sales, and marketing; and pricing. Companies should make sure they have a production

[12] For most financial exchanges the two sides are liquidity providers who announce prices at which they stand ready to buy or sell and liquidity takers who provide order flow that interacts with those standing orders. See Larry Harris, *Trading and Exchanges: Market Microstructure for Practitioners* (Oxford University Press, 2002). For non-financial exchanges such as e-commerce the two sides are simply buyers and sellers.

[13] *See*, e.g., Susan Hart and Nikolaos Tzokas, "New Product Launch "Mix" in Growth and Mature Product Markets," *Benchmarking: An International Journal*, Vol. 7, No. 5, (2000): 389-406 and Biren Prasad, "Analysis of Pricing Strategies for New Product Introduction," *Pricing Strategies and Practice*, Vol. 5, No. 4, (1997): 132-141. For literature summary see, V. Krishnan and Karl T. Ulrich, "Product Development Decisions: A Review of the Literature," *Management Science* 47, no. 1, (Jan., 2001): 1-21.

and distribution system for getting products and services to consumers. They need to provide sell consumers on the merits of trying a product through the dissemination of information as well as persuasion. Finally, they need to set prices recognizing both the competition and the fact that consumers may need an incentive to try to a new and unproven product. These traditional strategies facilitate obtaining consumers on both sides of a two-sided platform, that is, in making horizontal and vertical movements.

This section focuses on several aspects of product launch and product design that are likely to be particularly applicable to the ignition problem faced by catalysts. Before we begin it is useful to clarify some issues concerning the relationships among different platform members which will affect the discussion below.

For some platforms, customer groups are very distinct. Companies such as Electronic Arts that develop and sell videogames for the Sony PlayStation platform are distinct from people who buy and use Sony PlayStations.

For other platforms, the customers appear so similar that the platform may not appear multi-sided at all. A telephone platform connects people who talk to each other. People are people. Closer inspection, though, often reveals that people fall into one of two distinct positions on the platform at any point in time, that they need people in the other positions to connect to, and that it is possible to manipulate the pricing structure to mediate the externalities between these two groups.[14] At any point in time members of a phone network are either calling someone or being called and carriers can adopt pricing structures to alter the incentives to make or take calls. The same is true for social networking sites that involve inviting and accepting friends and then involve participating in other interactions that entail initiating or receiving messages communications.[15]

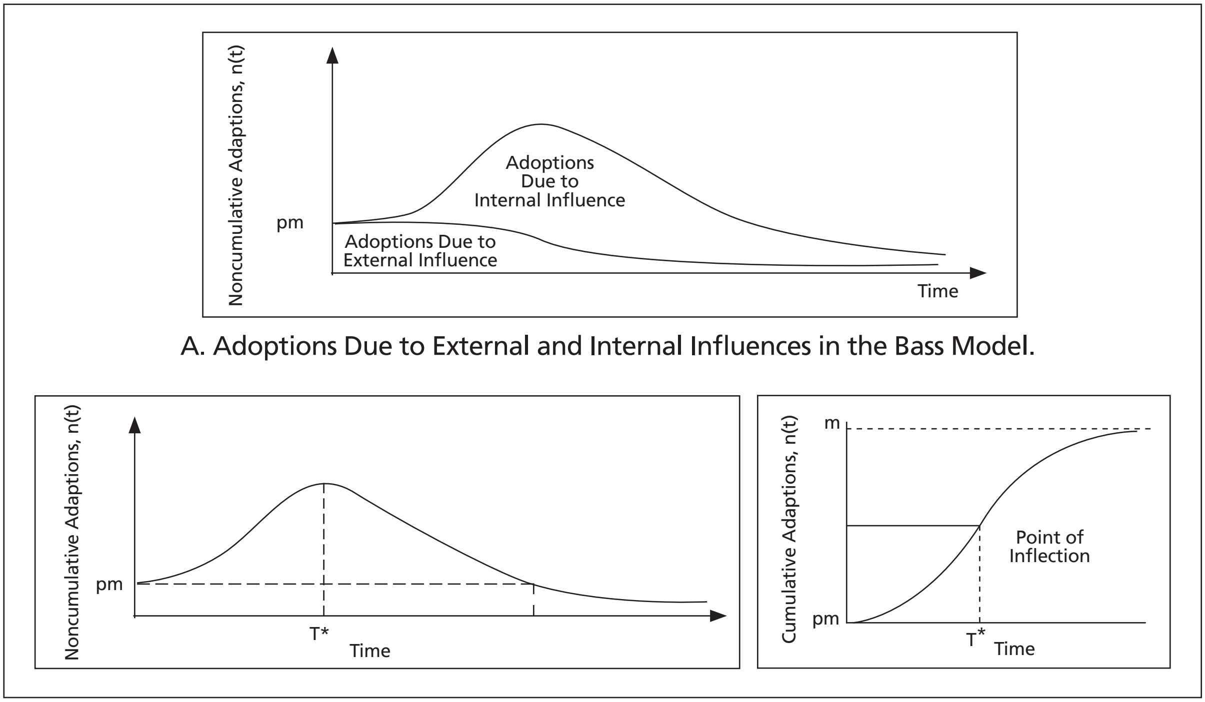
The fact that customers on the different sides may be the same economic agents clearly facilitates platform ignition. While the catalyst entrepreneur might have to provide incentives for these economic agents to engage in different types of behavior it can focus on securing the participation of one well-identified group of agents.

---

[14] The fact that the pricing structure—the relative charges to the two groups matters—makes these a two-sided platform under the definition proposed by J.C. Rochet and J. Tirole, "Platform Competition in Two-Sided Markets," *J. European Economic Association* 1, no.4 (2003): 990-1029. See e.g, Jerry Hausman, and Julian Wright, "Two Sided Markets with Substitution: Mobile Termination Revisited," 2006, mimeo. Using data from Australia they explore the equilibrium prices of fixed-to-mobile calls by taking into account that fixed-line callers might be cell-phone users themselves, and thus be from either side of the market. They also point out that it is important to consider whether the mobile receiver pays for the call (as in US) or does not (as in most EU countries).

[15] The social networking literature distinguishes between senders and receivers and highlights the important role that gregarious members of a network (i.e. individuals who initiate many interactions) and prestigious members of a network (i.e. individuals who are the recipients of many messages) play. *See* Wasserman, S and Katherine Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).

FIGURE 2 The Basic Model of Product Diffusion



Source: Mahajan, Muller, and Bass (1990)

A. Product Diffusion

The original models of product diffusion[16] distinguished between two types of consumers: innovators who would try a product as a result of direct communication with them and imitators who would try a product as a result of communication with someone else who had tried the product. The innovator might learn about the product through advertisements in the mass media. The imitator might learn about the product either from the innovator or from other people who learned about it directly or indirectly from the imitator. The word-of-mouth aspect to this model gives rise to the well-known S-curve of product diffusion where there is a convex rise in adoption, an inflection point, and a concave rise that levels off at some saturation point. The population of economic agents is sometimes broken down into innovators, early adopters, early majority, late majority, and laggards. Figure 2 shows the standard framework.[17]

[16] *See* Vijay Mahajan, Eitan Muller, and Frank M. Bass, "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing* 54, No. 1, (Jan., 1990): 1-26

[17] *See* Vijay Mahajan, Eitan Muller, and Frank M. Bass, "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing* 54, No. 1, (Jan., 1990): 1-26.

The literature on social networks provides insights into the process of word-of-mouth communication.[18] The social graph describes the relationships among members of a network. It consists of nodes which reflect the agents and lines that show the connection between the agents. The connections can be unidirectional (x communicates with y) or bidirectional (x and y communicate with each other). The nature of the relationships among members of the network can provide insights into the organic workings of the network. There are three key ramifications for product diffusion. Word of mouth will spread more quickly: (1) the more connections innovators have; (2) the more connections friends of the innovators have; and (3) the denser the network is in the sense of there being fewer degrees of separation among members of the network.

Two types of agents facilitate the diffusion information within the network. "Influencers"—or gregarious members—account for a disproportionate share of communications. They send a lot of messages out to a broad range of connections. "Centers" have connections with many people who are not connected to other agents.[19] They are important because they are the only way to reach isolated agents within the network.

Product diffusion may provide double duty for two-sided businesses in which economic agents have shifting roles in the platform. Individuals who upload photos on the photo-sharing network Flickr may view photos from their friends. People who use eBay's auto exchange may use it for both buying and selling, though some users may specialize in selling, and some may never sell.

B. Direct Network Effects

Consumers may value a product more if similar consumers use that product as well. This is known as a positive direct network effect.[20] It can arise because it is easier to connect to people using the product or because there are knowledge spillovers among them. Consumers may also value a product less if similar consumers use that product. This is known as a negative direct network effect. That might happen because of congestion or because people want to be different. For simplicity, we will assume that direct network effects are positive unless noted otherwise.

Direct network effects act as an accelerant to a catalytic reaction. Diffusion happens more quickly. The value of the network is higher to each additional member who is contacted. All else

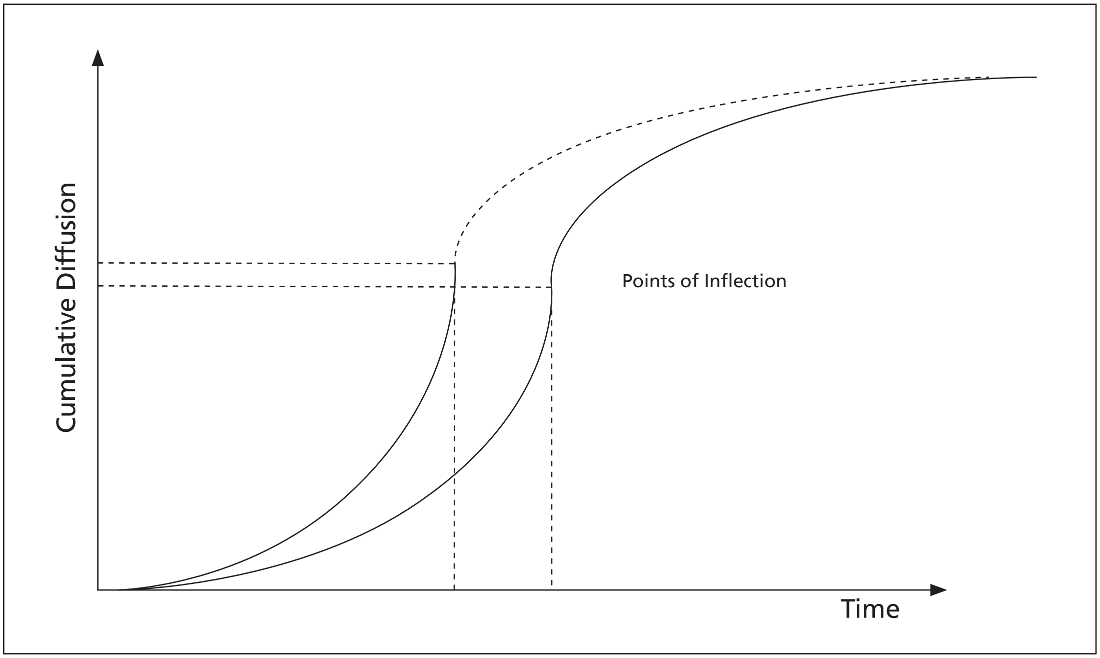
---

[18] *See* Jacqueline Brown and Peter H. Reingen, "Social Ties and Word-of-Mouth Referral Behavior," The Journal of Consumer Research,Vol.14, No. 3, (Dec. 1987): 350–362, and Jo Brown, A. Broderick, and Nick Lee, "Word of Mouth Communication Within Online Communities: Conceptualizing the Online Social Network," *Journal of Interactive Marketing* 21, no. 3, (2007): 2–20.

[19] The centers are the nodes for "stars" in which the center is connected to other nodes and the points of the star are connecter to the center node.

[20] *See* Carl Shapiro and Hal. Varian, *Information Rules, A Strategic Guide to Network Economics,* (Harvard Business School Press, 1998).

FIGURE 3 Network Effects and Diffusion



equal direct network effects increase the likelihood that each subsequent agent that is contacted by an earlier adopter will adopt the product also. Figure 3 shows how direct network effects modify the S-curve of diffusion.

Social networking theory provides some guidance on how to use direct network effects strategically. I would conjecture that economic agents that are more densely connected in a network have stronger direct network effects among them. Thus using influencers to connect to more densely connected portions of networks will tend to have higher payoffs.

C. Indirect Network Effects

One type of economic agent may value a product more if more of another group of economic agents uses that product as well. This is known as a positive indirect network effect.[21] It can arise because one type of economic agents (e.g. a buyer, a man, cardholder) wants to search for and transact with another type of economic agent (e.g., a seller, a woman, a merchant) and vice versa. It can also arise because one type of economic agent (e.g. a computer user, a video game user) wants to be able to find complementary products for the platform they use (e.g., applications,

[21] *See* Shapiro and Varian, *id.*

video games) and the maker of those products want to be focus its efforts on platforms that have users who will demand its products. There are also negative indirect network effects: one type of economic agent on the network harms another type of agent. The leading case of this for platform businesses is advertising-supported media. Consumers may dislike the advertisements. The platform solves the externality problem between advertisers and consumers by using content to bribe people into viewing ads.

Indirect network effects are the key aspect of multi-sided platforms. They are the source of the catalytic reaction—and much of the value—created by the platform. A key practical aspect of these indirect network effects is that they require that the platform "balance" the two sides to maximize the value of the platform to either side. The platform has zero value to either side if the other side is not on board. For many platforms the optimal balance is likely well into the interior and away from the axes as was shown in Figure 1.[22]

D. The Role of Customer Heterogeneity

The analysis above has already pointed to the fact that not all customers are created equal for multi-sided platforms. It is useful to pull these concepts together here and summarize their implications for platform ignition. There are three major kinds of heterogeneity.

First, some customers value a product or service more than other customers. All else equal those are the ones to go after initially to grow a business because they involve the lowest cost of sales and marketing. They can then kick off product diffusion. Two-sided platforms sometimes need to recruit these customers on both sides.

Second, some customers on one side are valued more by customers on the other side. The two-sided literature calls these "marquee" customers[23] while the social networking literature calls them "prestige" nodes—i.e. nodes that many people want to connect to and therefore receive many messages. All else equal marquee customers are the most important one to attract early on. They not only increase the value of the platform but also bring in more customers on the other side who help stimulate product diffusion on that side. Marquee customers may appear on one or both sides.

Third, some customers are more gregarious than others in the sense that they are more likely to influence other customers to join the platform. These "influencers" are important to attract early on because they will accelerate the vertical or horizontal growth of the platform. To ignite, platforms want to identify and recruit heavy influencers on both sides early on.

Prestige and influencer customers both generate significant direct or indirect externalities. It therefore often pays to subsidize their joining the platform.

---

22 See Evans and Schmalensee, "Failure to Launch," op. cit.

23 *See* J.C. Rochet and J. Tirole, "Platform Competition in Two-Sided Markets," *J. European Economic Association* 1, no.4 (2003): 990-1029.

III. STRATEGIES FOR IGNITING CATALYTIC REACTION BY SOLVING COORDINATION PROBLEMS

This section focuses on diagonal strategies for getting both sides on board in the right proportions. Catalyst entrepreneurs must ultimately solve a coordination problem to get both sets of economic agents to get on board their platforms. The dynamics of this coordination problem can vary considerably depending on the nature of the platform business.

Sequential entry. In some cases it is possible to get one group of agents on board over time and then make these agents available to the other group of agents later in time. That is the situation with advertising-supported media. One can use content to attract viewers and then bring advertisers on board later. This dynamic works because there are non-positive indirect network effects between the two sides: viewers do not care about advertisers (and may dislike advertising) but come to platform for the content.

Entry with significant pre-commitment investment. In other cases one group of economic agents need to make investments over time to participate in the platform.[24] That is the case with software-based platforms such as video game consoles. Game developers must invest in creating games for the next release of a console without knowing how many consumers will be interested in using that platform when their development is done. The video-game console platform must either convince game developers that buyers will show up, provide them with some financial guarantees that buyers will show up, or self-produce games until the platform has demonstrated itself.

Simultaneous entry of sides. In some cases the economic agents are making decisions to join the platform around the same time and have to both join around the same time for the platform to provide value. A dating venue demands almost perfect simultaneity. Heterosexual men would quickly leave a new nightclub that had no women and vice versa. Other platforms provide more latitude. Buyers may not desert an exchange platform right away if there are no sellers but they will arrive soon.

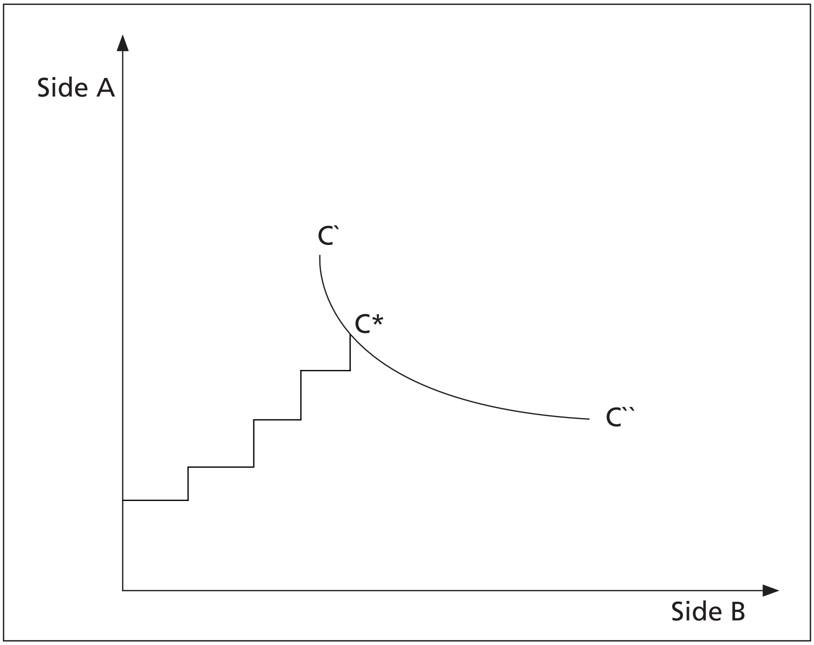
In all cases, however, platform growth is not sustainable until the platform reaches critical mass. Therefore the key challenge for new platforms is figuring out ways to reach critical mass quickly.

A. The Basic Zig-Zag

A basic strategy for reaching critical mass is to build participation on the two sides incrementally. The platform starts with a small number of economic agents on both sides. It then persuades agents on either side to join. It also relies on the natural processes of product diffusion. Because of indirect

[24] This is the case treated in Andrei Hagiu, "Pricing and Commitment by Two-Sided Platforms," *The RAND Journal of Economics 37, no. 3 (fall 2006)*: 720-737.

FIGURE 4 Basic Zig-Zag to Critical Mass



network effects, the platform is more valuable to each successive group of prospective customers. Figure 4 shows that basic zig-zag approach to growth.

eBilleMe provides an example of this strategy. Consumers who click on the eBillMe sign at the checkout for an e-tailer can pay with their online banking account. They then send an email which contains details for paying from their online banking account. After they enter the information into their online banking account they receive a receipt and the product is shipped. This payment alternative is attractive to people who are either concerned about the security of paying with cards online or do not happen to have a card conveniently available. A small but significant fraction of consumers, as it turns out, like paying this way.

To get started, eBilleMe persuaded ToolKing to offer eBillMe at checkout.[25] A small percentage of customers used this payment alternative. eBillMe then went to other online retailers. Each led to eBillMe having more people who were accustomed to using its service. For each subsequent merchant it went to it offered an increasingly valuable offer since it had more users who were predisposed to use this payment alternative. At the same time it let its users know that they could pay at more places thereby increasing the value to the merchants. eBilleMe grew from 1 merchant and hundreds of users during its first year of entry in 2005 to hundreds of online stores taking 2% to 10% of the merchant's transaction volume by 2008.

[25] Interview with Marwan Forzley, founder and chairman of eBillMe.

B. Pre-commitment to Both Sides

Some platforms such as eBillMe are able to start with one member on one side that it uses to attract members on the other side. More commonly platforms need to have multiple members of both sides to begin the zig-zag process above. They therefore need to persuade a minimum number of early adopters on both sides to show up at the start of the platform to make it credible. That requires getting both sides to believe that when the platform opens for business there will be members of the other side present.

Diners Club is the classic example of this strategy.[26] Although the precise sequence is lost to history it persuaded 14 restaurants in Manhattan to accept Diners Club cards for payment. At the same time it persuaded several hundred people in Manhattan to take the Diners Club Card. Those commitments were enough to start the platform. Over the next year Diners Club zig-zagged its way to 330 restaurants and 42,000 cardholders.[27]

Customers may require more assurance that the other side will in fact show up especially if they have to invest resources to join. Contingent contracts can be entered into for this purpose. Customers agree to commit to join the platform conditional on other customers on the same and other side also joining. Once these contracts have been entered into the catalyst only needs to persuade one customer to sign on because that will have a domino effect on all the other customers. MobiTV[28] serves as an example where contingent contracts play an important role for "catalyzing" the platform. Started in 1999, MobiTV's planned to offer TV service to customers on their mobile phones for a subscription fee. It needed to persuade TV content providers and mobile operators to join its platform. Neither of the two wanted to embark on the new project unless they were assured that the other side would also join. MobiTV used contingent contracts to ignite the platform. It signed an agreement with Sprint that if television channels join, Sprint would offer service and agreements with MSNBC and other TV channels that if Sprint join they would broadcast. It was enough for MobiTV to get Sprint to agree and everyone else followed.

The video-console example considered by Hagiu is an extreme example of this phenomenon.[29] Video-games and other software application are platforms that connect application developers and video game players. The game player will not purchase a video console without enough applications and games, and the former will not put the time to develop such if they are not sure that people will buy their applications. Because game developing is a long process, the vendors have to secure sellers a long time before the new game is launched in order to make sure that developers will provide the applications.

[26] For description and discussion of this example, see David S. Evans & Richard Schmalensee, *Catalyst Code*: *The Strategies Behind the World's Most Dynamic Companies* (Massachusetts: Harvard Business School Press, 2007), p. 1.

[27] At this point in time credit cards were mainly used in local markets so Diners Club had to secure critical mass in a number of different cities to ignite the platform. The 330 restaurants and 42,000 cardholders refers to the total across these separate markets.

[28] Id., p. 89.

[29] Hagiu, *id.*

C. Single and Double-Marquee Strategies

The marquee strategy discussed above is another way to obtain enough members on both sides to begin the zigzag to critical mass. In a single-sided marquee strategy the platform acquires an "influential" or "prestige" member of one side. Announcement of that may attract enough members of the other side at the beginning. The shopping mall strategy is the classic: the mall gets an anchor tenant which many shoppers want to connect to. In a two-sided marquee strategy the platform acquires "influential" or "prestige" members on both sides. They provide value to each other as well as attract other members. Nightclubs are common users of this strategy. They try to get popular men and women to come on opening night. They want to connect to each other and less popular men and women want to connect with them.[30]

D. The Two Step

The two-step strategy involves getting enough members of one side on board first and then getting members of the other side on board. As mentioned earlier this works when the first side does not value access to the second side which is often the case for advertising-supported media. Search engines followed this strategy. They attracted users who did searches of the world-wide-web. The search results were displayed on a series of pages. Once they obtained enough page views they sold access to those pages to advertisers. Google, for example, operated its search engine for 23 months (including a beta version) before it opened it search results pages to advertisers[31]. At that time it had more than a billion pages indexed and 18 million user queries per day.

E. Ziz-zag with Self-Supply

Catalysts may be able to jumpstart their platforms by providing one of the sides themselves at least initially. Consider YouTube which is a three-sided platform: user-generated content attracts viewers, viewers attract content providers who want an audience, and access to be viewers can then be sold to advertisers. YouTube started by focusing on users and viewers. Its founders seeded the site with content they generated themselves and started the process of diffusion by suggesting that members of their personal social networks check out the content.[32] They also used various marketing strategies to attract viewers: they posted an ad on craigslist to compensate attractive women to post on the site and promised to give an iPod to a random user every day till the end of the year.

---

[30] See Evans & Schmalensee, Catalyst Code, id.
[31] Google Milestones: http://www.google.com/corporate/history.html.
[32] See YouTube videos: "The History of YouTube" and "The Real History of YouTube in 3 minutes."

F. Summary

We have introduced several concepts.

- First, multi-sided platforms often must attain critical mass to ignite a catalytic reaction that leads to organic growth. Platforms that do not reach this critical mass implode.
- Second, to reach critical mass platforms can engage in a number of strategies to get "enough" customers on either side, and in the right proportions, on board. These include the zig-zag and the two-step.
- Third, these strategies can usefully employ many of the tactics used for new product introductions by non-platform businesses.

We have taken some liberties in explaining the dynamics of platform ignition. As a formal mathematical matter one would have to reach critical mass instantaneously to ignite the reaction. In practice it appears that platforms have some limited time to get to critical mass. Early adopters use a platform. If they come back and if later adopters also find value then it is possible to reach critical mass. If the platform does not grow quickly enough to critical mass early adopters lose interest, fewer later adopters come, and word-of-mouth referrals stop or turn negative.

IV. B2B EXCHANGES

Entrepreneurs and investors flocked to developing B2B exchanges in the late 1990s.[33] The thesis was simple. Build a better platform for exchange and buyers and sellers will flock to it. The failure was stark. When the exchanges opened their doors few sellers showed up. The sellers largely kept staying away.[34] Buyers lost interest. The exchanges could not reach critical mass. They failed en masse in the 2001 dot.com bust.[35] Despite the great promise many held out for this new way of doing business few successful B2Bs have emerged in the post 2004 web boom.

The B2B platforms failed to ignite for three related reasons.

First, a new platform must create a significant value as a result of getting two sides together that exceeds what they can do on their own or on alternative platforms. The B2Bs did not provide enough value compared with buyers and sellers connecting without a platform through existing bilateral relationships or an existing offline consortium.

[33] There were over 1600 B2B exchanges by 2001 according to David Lucking-Reiley and Daniel F. Spulber, "Business-to-Business Electronic Commerce," *Journal of Economic Perspectives 15, no.1 (2001):55-68*.

[34] That was the story of Chemdex, one of the first B2B exchanges. Chemdex did not manage to attract enough sellers and buyers in order to secure the authomation of the business process. *See* Flora Nguyen, "The Changing Face of E-Healthcare, Is healthcare too complicated for an Internet-based supply-chain solution? Here's what manufacturers should expect in 2001," http://www.devicelink.com/mx/archive/01/03/0103mx072.html

[35] See, e.g., Jack Nickelson and Hideo Owen, "A Theory of B2B Exchanges Formation," 2001, manuscript.

Second, a new platform must allocate that value between the two sides to provide both with a sufficient incentive to join the platform. The B2Bs were not able to offer enough of a value proposition to sellers to get enough of them on board. The sellers saw the B2B exchanges as methods for driving down their prices through auctions. That did not leave them with enough of a long-term return and they therefore stayed away.

Third, a new platform must get enough of both sides on board and in the right proportions to achieve the critical mass that provides a minimum amount of value to both sides. Sellers came on board too slowly. Buyers lost interest. That provided even less motivation for sellers. Platforms that do not achieve critical mass implode and never have the chance to grow.

A. The Rise and Fall of the B2B Exchanges

B2Bs were viewed as an obvious way in which the Internet could create great value. In the late 1990s and early 2000s business strategists, economists, and others wrote numerous articles on how B2Bs would transform buying and selling. The title of one of the business books on the subject provides a flavor: *B2B Exchanges: The Killer Application in the Business-to-Business Internet Revolution*.[36] Various researchers forecasted that B2Bs would become account for a large fraction of commerce. Goldman Sachs predicted in 2000 that B2B e-commerce transactions would equal $4.5 trillion worldwide by 2005. The Gardner Group estimated in 1999 that by 2004 B2B e-commerce would reach $7.3 trillion.[37] Entrepreneurs and venture capitalists poured into this new industry. Between 1995 and 2001 there were more than 1,500 hundred B2B sites. Some of the most prominent ones were Ventro, VerticalNet, Neoforma, Cordiem, CorProcure, Chemdex. Most of them collapsed in the early 2000s as investors realized that they did not have a viable business model and as the expected buyers and sellers failed to turn up.[38] Many of the ones that survived—such as Ariba, which merged with FreeMarkets—shifted their focus from operating true exchanges to offering procurement software that facilitated the normal process of bilateral buyer-seller transactions.

Perhaps the best evidence on what happened to not only B2Bs, but the idea of B2Bs, is revealed from a simple Google search. News and analyst reports about B2Bs largely end in the early 2000s. The flourishing academic literature on the economics of business of B2Bs appears to have collapsed soon thereafter. Surprisingly, there are few rigorous port-mortems on why B2Bs failed to ignite. This chapter does not provide a rigorous study either but places some of the contemporaneous observations about the collapse in the context of the framework presented above.

---

[36] Arthur Sculley and William Woods*, B2B Exchanges: The Killer application in the Business-to-Business Internet Revolution* (ISI Publications, 1999).

[37] David Lucking-Reiley and Daniel F. Spulber,"Business-to-Business Electronic Commerce," *Journal of Economic Perspectives* 15, no. 1, (2001): 55-68.

[38] According to a research study, *Shakeouts in Digital Markets: Lessons from B2B Exchanges*, conducted by Wharton School and Pembroke Consulting in November 2002, of the 1,500 independent business exchanges that existed in 2001, less than 200 remained at the end of 2003.

B. The Value Proposition

There were well developed methods that enabled business buyers and suppliers to deal with each other well before the development of the commercial Internet. Buyers had supplier lists which they would turn to when they had a need. They had sophisticated procurement departments that experience in putting the purchase of goods and services out to bid and had knowledge about suppliers. Suppliers also had relationships with companies and had ways to get on the bidding lists. Buyers had the greatest difficulty finding new suppliers especially for areas where they had few bidders. New suppliers who had not developed relationships also had difficulty. A key question for starting a platform was whether it could provide enough additional value to these two market sides.

The B2Bs were patterned after the successful auction-based B2C sites such as eBay. They introduced various types of auction mechanisms that tended to maximize competition among suppliers but mainly based on price. That of course was not attractive to suppliers because it depressed prices and profits and eliminated the value of other sources of differentiation. As one consultant noted about the proposed B2B exchanges for the airline industry:[39]

> [suppliers] continue to be reluctant to sign up to portents and other e-mechanisms created by the prime contractors. The key reason for this is that the primary objective of e-procurement is perceived to be a reduction in the purchase price, therefore forcing pressures on [supplier] margins.

A number of observers of the demise of the B2Bs observed that a major problem was that suppliers were scared, as Kabir puts its, "of comparison shopping and brand dilution."[40]

Once we go beyond the perfectly competitive model it is easy to see why exchanges that were built around efficient procurement auctions were not attractive to suppliers. Most businesses have fixed costs that they have to recover. They therefore cannot survive if their prices are competed down to short-run variable costs. Department stores can have sales periodically where merchandise is sold at low margins; they could not survive if they were selling at those prices all the time. B2B auctions will tend to attract suppliers that can offer low prices because they have excess inventories as a result of lack of success.

The B2Bs were not necessarily advantageous to buyers either. Company procurement officers typically learn about the quality of their suppliers. Through interactions over time and discussions with people in the industry they discover the quality of supplier goods and services as well as their reliability for delivery. They can take this information into account in deciding who to put on the bid list thereby selecting the lowest price from pre-qualified bidders, or they can take non-price

[39] *See*, Mark Odell, "B2B Struggle to Achieve Take-off," *Financial Times*, June 18, 2001.
[40] *See* Nowshade Kabir, "E-marketplace; Facts and Fictions," WebProNews, posted on August 20th, 2003, http://www.webpronews.com/topnews/2003/08/20/emarketplace-facts-and-fictions.

features into account in selecting the winning bid. Many of the B2B exchanges focused mainly on price and did not provide companies with significant services for assessing quality. Indeed, the importance of quality and experience was learned the hard way by American toy manufacturers who found suppliers through some of the successful B2Bs that have focused on connecting Chinese suppliers to multinational buyers. Many of these suppliers used lead in their products which led to significant health risks and product recalls when this was discovered.[41]

C. Ignition Strategies

Based on the general articles that have been published on the mass demise of the B2Bs and several B2Bs it appears that the B2Bs made several fundamental mistakes in achieving critical mass and in igniting. To begin with it is not clear that many of these B2Bs could have ignited without modifying their business models to provide sufficient value to the two sides of the platform. That is, it is not clear that the global communication aspect of many of their models, based on the Internet, provided enough value over and above the traditional methods of procurement to warrant buyers or suppliers to change. Therefore, there may not have been a pricing structure, and ignition strategy, that could have, under any circumstances, sustained many of these B2Bs.

Some of the B2Bs seem to have adopted the following two-step strategy. They would begin by organizing the buyer side. That could occur by securing the participation of one or more large buyers—a one-sided marquee strategy. For example, Chemdex started with Genentech which purchased most of the value on the exchange initially. That could also occur as a result of the exchange being started by a cooperative of buyers—that was the case with the automobile exchange, Covisint, which was started by General Motors, Ford Motor Company, DaimlerChrysler, Nissan, and Renault. They could then seek suppliers. This strategy would not work for exchanges that were organizing larger groups of smaller buyers. They would have to adopt a simultaneous strategy and secure enough buyers and sellers at the same time.

Many of the B2Bs made the mistake of establishing symmetric pricing structures that sought to earn revenue from both buyers and suppliers.[42] One of the most famous B2B failures—Chemdex—charged suppliers a listing fee for products and charged buyers a commission on transactions.[43] It is possible that a symmetric pricing strategy could work if an exchange had enough

---

41 David Barbosa, "Mattel recalls toys made in China," *Herald Tribune*, August 2, 2007.

42 See the case studies in Anne Engstrom and Esmail Salehi-Sangari, "Assessment of Business-to-Business e-Marketplaces' Performance," (PhD diss., Luleå University of Technology, 2007). Of them the 6 B2Bs they examined, five started trying to earn revenues from both buyers and suppliers and one of these moved towards a model in which suppliers obtain services largely for free.

43 See M. Meyer, Neil de Crescenzo, and Bruce Russell, "Case Study: Chemdex: In Search of a Viable Business Model," *International Journal for Entrepreneurship Education* 2, no.2. (2004), http://web.cba.neu.edu/~mmeyer/cases/Ventro-010105.pdf

buyers and suppliers already on—at that point suppliers might lack access to a significant market if they refused to pay the fees. But at the start of these exchanges buyers and sellers would have to be persuaded to switch from the widely used bilateral procurement methods. In any event, it appears that many of the exchanges ended up well off the path to critical mass because they did not offer suppliers with enough of an incentive to join. That could have come in the form of subsidies in cash or kind to suppliers to join the exchange. If the B2Bs could have gotten enough suppliers through what is known as a "divide and conquer" strategy they might have gotten enough buyers and then more suppliers.[44] Of course, such subsidies could have been large risky investments in the success of the B2Bs and granting them would have been foolish if, as suggested above, there was no feasible profitable model for the B2B.

V. SOCIAL NETWORKING

Social networks have exploded on the web. By August 2008 there were more than 110 social networking sites, as classified by Alexa[45], which had more than 580 million active users.[46] In the United States MySpace is the largest one with more than 68 million active users. The second largest one, Facebook, has about 32 million active users.[47] These sites generally allow people to construct a public profile, make this profile available to other users to whom they grant permission, and in return obtain the right to see the profiles of these users. The profiles and related technologies provide ways for friends to communicate with each other. For example, Facebook has a place where people can say what they are doing; when a user goes to her profile page on Facebook she will then see what all of her friends are doing.

The first social networking site, SixDegrees.com, was started in 1997. But it did not attain critical mass and shut down in 2000. Its founder speculated that was because people did not have enough friends online and there was not much to do together once they were on line. The promise of social networking became apparent with the rapid growth of Friendster which started in 2002. In the United States, it grew to a peak of 1.5 million users in the third quarter of 2003 in the United but then went into decline in the United States. MySpace and Facebook have emerged as the two leading global social networking platforms.[48]

---

[44] *See* Jack Nickelson and Hideo Owen, "A Theory of B2B Exchanges Formation," 2001, manuscript.

[45] http://www.easy-viral-traffic.com/blog/index.php/top-list-of-social-networking-sites/

[46] http://www.comscore.com/press/release.asp?press=2396

[47] ComScore Data

[48] See D. Boyd and N. Ellison, "Social Networking Sites: Definition, History, and Scholarship," Journal of Computer-Mediated Communication, 13:1, 2007, available at http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html, last visited on September 27, 2008. In addition to these "friending" sites there are also professional social networking sites such as Linked In which in theory help people form professional relationships and in practice help people find jobs and recruiters to find candidates.

A. Background on Social Networking

Social networking sites are multi-sided platforms with usually at least three sides.

The first two sides are individuals who want to connect with each other. For many social networking sites such as Friendster these are friends seeking friends. However, people seek to find friends on social networking sites and people are recipients of requests for friendships. This sender-receiver relationship has important consequences for how social networking sites are ignited. The role of gregarious and influential members of the site can have significant effects on the dynamics of growth. The third side consists of advertisers who want to reach these users. Some social network sites a fourth side that consists of software developers who write applications that work with the social networking site and provide value to users and advertiser.

Social networking sites typically have two sources of revenue. In part they are traditional advertising supported media. The social networking features and applications attract viewers. Advertisers then pay for the ability to present advertising messages to these viewers. The social networking site earns revenue directly from advertisers or indirectly from application providers who sell advertising. They also earn revenue by selling things that facilitate interactions between the users. These include "pokes" such as virtual flowers that people can send their friends.

Social networking sites use a variant of a two-step to ignite. They focus on getting a critical mass of "friends" which generates traffic on their sites. They then earn revenue through selling advertising and pokes. They also persuade developers who want access to their network to write applications which further increase the value of the network as well as provide for additional revenue possibilities.

We now turn to the ignition strategies and growth paths for two well-known social networking sites: Friendster and Facebook. Figure 5 shows membership on these sites and the growth rates in membership over time.

B. Friendster

Friendster was the first social networking site to gain wide popularity.[49] It attracted users initially by providing a dating venue where people could find romantic partners among friends of friends. That was a different approach than existing online data venues that matched up strangers based on predictions of their likely compatibility. People posted their profile and were able to communicate with other users who were no more than four-degrees of separation away. People interested in finding romantic partners liked the fact that there were social connections which not only helped screen people but also helped police bad behavior.

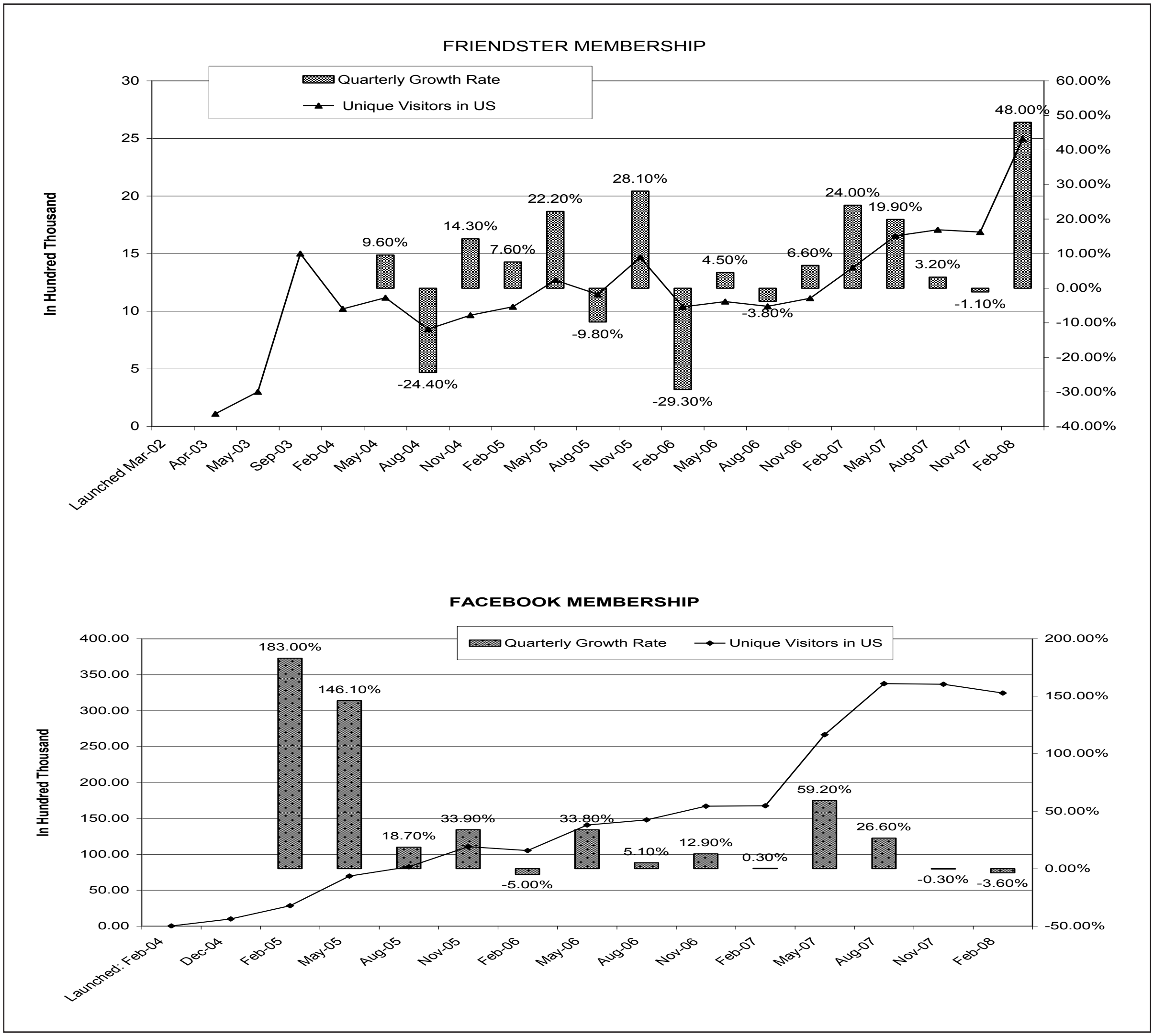
Friendster started with a prototype in 2002. Its founder invited 20 close friends to join.[50] The site launched formally in March 2003. It grew virally from the initial seed of 20 friends

---

[49] See Boyd, D. M., & Ellison, N. B., "Social network sites: Definition, history, and scholarship" *Journal of Computer-Mediated Communication, 13*(1), article 11 (2007).

[50] http://www.inc.com/magazine/20070601/features-how-to-kill-a-great-idea_pagen_2.html

FIGURE 5 Friendster and Facebook Growth



Note: Based on ComScore data for the period Feb 04 –Feb 08 and "Facebook Case Study", Harvard Business School, 2006, for the period Apr 03-Feb-03.

and had 835,000 registered users by June 2003 and 1.5 million by September. There is no evidence that it focused on influencers or gregarious people—instead it just grew naturally. The growth was not driven entirely by the users looking for people to date. Most people used it to communicate with and expand their network of friends. Many people used it as a source of entertainment from looking the profiles of people that they would not otherwise have contact with.

Although it reached critical mass and grew explosively Friendster encountered two major problems.[51] The first was technological. It was not prepared for the upsurge in users and users started finding that the site was not reliable. The four- degree of separation rule was a major source of the problem. Every time someone signed on to the site the software had to determine who was within the individual's community. As the size of the overall network grew it took longer and longer to do that calculation.

The second involved changes in the community and Friendster's response to those changes. As the community grew the four-degree of separation rule made people's profiles accessible to people that they did not necessarily want in their social network—for example their bosses. At the same time the four-degree of separation rule limited the size of the community that one could acquire. To counter this, people started gaming the system to create more friends. One of these efforts involved creating Fakesters that were fake profiles that were designed to attract a lot of attention and therefore many friends. Friendster banned fake profiles and also dropped the "most popular" attribute that had encouraged people to engage in some of the bad behavior. As it turns out, the Fakesters were a popular source of entertainment and way to create groups such as alumni of particular universities. Friendster also policed the profiles and kicked users off who did not comply.

The user community became dissatisfied with Friendster. Growth slowed.[52] In the U.S. market which is the most significant source of advertising revenue Friendster declined dramatically after the third quarter of 2003 and remained roughly flat for the next three years while other social networking sites grew explosively. Nevertheless, it achieved a critical mass of users in other countries, and remains the 3rd largest social networking site worldwide;[53] in August 2008, announced that they received a $20 million of venture capital.[54]

C. Facebook

Facebook's founders also started with the idea that they would provide a convenient method for dating. As students at Harvard they thought it would be helpful to have a place where students who saw each other in class and other settings could connect. Within the first month of starting in February 2004, more than fifty percent of Harvard college students had put their profiles on

[51] *See* Gary Rivlin, "Wallflower at the Web Party," New York Times, October 15, 2006.

[52] MySpace picked up a number of Friendster's users. It provided a platform in which people could see anyone's profile and in which people had a great deal of freedom in designing their profiles. An important group that it attracted were indie rock groups, some of whom had been ejected from Friendster, and then musicians more broadly who used the site to connect with their fans.

[53] http://www.prnewswire.com/cgi-bin/stories.pl?ACCT=109&STORY=/www/story/08-18-2008/0004869579&EDATE=

[54] See Tom Taulli, "Friendster is still alive and well ... and gets $20 million," Bloggingstocks, posted on Aug 5, 2008, http:// ww.bloggingstocks.com/2008/08/05/friendster-is-still-alive-and-well-and-gets-20-million/ (accessed September 24, 2008).

the site.[55] Since Facebook was designed to facilitate communication within a closed community it was able to obtain critical mass—enough people found it valuable enough to use the site—very quickly and then captured most of the population of potential users.

Shortly thereafter Facebook extended its model to other universities beginning with Stanford, Columbia, and Yale. In each case a significant portion of the undergraduate student body set up profiles. By the end of the year Facebook had almost 1 million active users. All users had to have an ".edu" email extension to verify that they were part of an academic community. Facebook continued to add local networks of students. It had more than 800 college networks by May 2005. It then added international schools and work environments. By August 2006 it had 14,782 million registered users.

Facebook opened its doors to everyone in September 2006. However, it offered a much different level of privacy than either the imploding Friendster or the exploding MySpace. When a person creates a profile on Facebook you cannot interact with anyone else at the beginning. That person can build a community of friends either by sending out friend requests, or by being invited by others to be a friend. Users can also "de-friend" people and have various ways to restrict access to portions of their pages.

Facebook did not have any significant revenue stream for its first 36 months and focused instead on building its user base. In February 2007 when it had 16,737 million unique users worldwide[56] it introduced Facebook Gifts. These are icons such as flowers, hearts, and balloons that users can purchase for $1 can send to other friends. In May 2007 they launched Marketplace for classified listings for sales, housing, and jobs. In November 2007 the company launched Facebook Ads, an ad system that allows businesses to target their advertising to the precise audience they seek to reach[57]. As of September 2008, it is estimated that Facebook earns $350 million[58] from advertising and $35 million[59] from the sale of gifts.

Facebook's successful ignition was based on a form of the two-step. The first step involved developing the user base. There was a chicken-and-egg problem to be solved there since the network could grow only if there was a reciprocal relationship between enough people. However, by providing an effective vehicle for communication social networking sites can grow very quickly virally as friends invite friends who invite more friends.

The second step involved earning revenue from that user base. An important aspect of that was selling access to the user base to advertisers. Facebook and other social networking sites have been disappointed at their ability to secure advertising revenue. Online advertising is sold on the basis

---

[55] http://www.guardian.co.uk/technology/2007/jul/25/media.newmedia
[56] http://www.comscore.com/press/release.asp?press=1519
[57] http://www.facebook.com/press/info.php?timeline
[58] http://www.emarketer.com/Article.aspx?id=1005257
[59] http://lsvp.wordpress.com/2008/01/23/facebook-digital-gifts-worth-around-15myear/
http://facereviews.com/2008/09/02/facebook-virtual-gifts-make-big-bucks/

of "cost per thousand" viewers of an advertisement. The average CPM for online advertising is in the range of $10-15. The average for Facebook is reportedly below $0.50. (Of course although Facebook may not prove as valuable as its founders and investors might have liked it is likely to be enormously profitable.)

Part of the difficulty which appears not to have been anticipated is that advertising is a bit alien to a social networking community. It is natural to see advertising on publisher web sites. It is less natural to see advertising on your own profile page or the profile pages of friends—especially if you have been using Facebook for a while without such advertising. Facebook alienated its customers when as part of its Beacon advertising strategy it started tracking users` purchases from third-party websites. Facebook and other social networking sites are looking for innovative ways to advertise that satisfy the needs of advertisers without reducing the value of the social networking aspects of the sites.

The social networking sites illustrate the power of diffusion in generating web traffic. The successful sites grew very quickly because they provided existing physical networks of friends much more powerful tools for interacting with each other more efficiently virtually. That, however, is only the first step in a sequential entry strategy. The later strategies have required getting advertisers on board the platform.[60]

VI. CONCLUDING REMARKS

In this chapter I have tried to raise some of the important issues and provide some preliminary thoughts on them. There is considerable room for further theoretical and empirical research. On the theoretical front we need dynamic models that address the creation of multi-sided platforms especially in the case in which the platform needs to get both sides on board at the same time. On the empirical front we need research on the successful and unsuccessful launch of platforms. Both econometric studies of platform dynamics and case studies would be informative.

[60] At this point in time the social networking sites have, as noted above, found it difficult to obtain much interest from advertisers and are charging vastly lower prices for viewers than are other content-oriented sites.

Part Two

TWO-SIDED ANTITRUST ECONOMICS

CHAPTER

4 Antitrust Economics of Two-Sided Markets

David S. Evans

ABSTRACT

"Two-sided" markets have two different groups of customers that businesses have to get on board to succeed—there is a "chicken-and-egg" problem that needs to be solved. These industries range from dating clubs (men and women), to video game consoles (game developers and users), to credit cards (cardholders and merchants), and to operating system software (application developers and users). They include some of the most important industries in the economy.

Two-sided firms behave in ways that seem surprising from the vantage point of traditional industries, but in ways that seem like plain common sense once one understands the business problems they must solve. Prices do not and prices cannot follow marginal costs in each side of the market. Price levels, price structures, and investment strategies must optimize output by harvesting the indirect network effects available on both sides. By doing so, businesses in two-sided industries get both sides on board and solve the chicken-and-egg problem. There is no basis for asking regulators or antitrust enforcers to steer clear of these industries or to spend extra effort on them. The antitrust analysis of these industries, however should heed the economic principles that govern pricing and investment decisions in these industries.

I. INTRODUCTION

Dating clubs—typically bars or cafes—are an innovative way for men and women to meet each other in Japan.[1] At one club, men and women sit on opposite sides of a glass divide. If a man sees a woman he likes, he can ask a waiter to carry a "love note" to her. Dating clubs sell patrons the prospect of making a match. Their business works only if they attract enough members of the opposite sex to their club to make a match likely. Enough men must participate to attract women, and enough women to attract the men. The club must figure out how much to charge men and

[1] Howard W. French, *Osaka Journal; Japanese Date Clubs Take the Muss Out of Mating*, N.Y. TIMES, February 13, 2000.

women so that the club gets the right number and mix of patrons, while at the same time making money since most of these clubs are in business to turn a profit. One bar does this by charging men $100 for membership, plus $20 a visit, and letting women in for free.[2] The bar presumably believes that at equal prices it would attract too many men, or too few women. That pricing structure—one that obtains a disproportionate share of the revenues from the men—seems common, based on an unscientific survey, in singles bars, discotheques and other legitimate businesses around that world that help men and women find companionship.[3]

Dating clubs are an example of a "two-sided market."[4] In such a market there are two sets of customers who, in effect, need each other. Each type of customer values the service more if the other type of customer also buys the service. Businesses service such markets by acting as "matchmakers." To do so, they must get both types of customers on board to have a product to sell. Indeed, in such markets the product or service is consumed jointly by two customers and, in a sense, only exists at all if a "transaction" takes place between them.[5] For example, the "product" created by the dating club is, in the end, a date from which both parties expect to benefit, and which occurs only if both of them participate. A fundamental economic characteristic of two-sided markets is the presence of positive externalities from having the other side on board (lots of guys to meet) and the inability of the parties to the transaction to internalize these externalities themselves

---

[2] *Id.*

[3] Here are some examples based on recent (Web-site) visits: C2K, a dance club in Las Vegas, is free for local women while the cover charge is $10 for out-of-state women and $15 for men (visited Aug. 15, 2002) <http://www.lasvegas.com/localmusic/danceclubs.html>; the Buddha Lounge in Chicago charges $5-$15 less to women, depending on the day of the week, than to men (visited Aug. 15, 2002) <http://centerstage.net/dance/clubs/buddha-lounge.html>; and on Saturday nights, The Wave Nightclub in Atlantic City lets women in for free while men are assessed a cover charge of $10 (visited Aug. 15, 2002) http://www.poolspanews.com/2002/011/ac_nightlife.html>. A recently developed online matching service has chosen equal prices—it specializes in matching identical twins. Twins Seek Twins in Online Matchmaking First, REUTERS NEWS, Apr. 11, 2002; and Twins Realm website (visited Aug. 15, 2002) <http://www.twinsrealm.com/>.

[4] The general economics of two-sided markets are discussed in a seminal paper by Jean-Charles Rochet & Jean Tirole. See Jean-Charles Rochet & Jean Tirole, Platform Competition in Two-Sided Markets, Financial Markets Group Discussion Paper #dp0409, (Nov. 26, 2001) <http://www.idei.asso.fr/Commun/Articles/Tirole/PlatformNov26.pdf>. *See also* Bernard Caillaud & Bruno Jullien, *Chicken & Egg: Competing Matchmakers,* Cepr Working Paper #2885 (Apr. 24, 2001), at 5; and Bruno Jullien, *Competing in Network Industries: Divide and Conquer,* Idei Working Paper, (Jul. 2001). Many of the notions discussed in this article were first introduced in papers that analyzed the payment card industry as a two-sided market. See, e.g., Richard Schmalensee, *Payment Systems and Interchange Fees*, L(2) J. Indus. Econ. 103 (Jun. 2002); and Jean-Charles Rochet & Jean Tirole, *Cooperation Among Competitors: Some Economics of Payment Card Associations,* Cepr Working Paper #DP2101 (Mar. 1999). This work is based in part on notions that were first recognized in W. F. Baxter, *Bank Interchange of Transactional Paper: Legal and Economic Perspectives,* 23(3) J.L. & Econ. 541 (1983).

[5] Although we shall see shortly that in many two-sided markets the parties do not even know one another. There is an intermediary who matches them and each only deals with that intermediary.

(only one or a few of the guys get your attention).[6] Firms profit themselves and society by figuring out ways to internalize these externalities.

Many important industries, including some central to what is often called the "new" economy,[7] are based on business models that are similar to dating clubs. Computer operating systems provide features that software developers can use in writing applications, and a platform on which computer users can run applications that use those features. Both developers and users must use the operating system for it to be a viable product. The PalmOS for handheld devices, Microsoft Windows for the desktop, and Sun Solaris for servers all depend for their success on obtaining application users and application developers. Videogames are similar. People who buy video game consoles, such as the Sony PlayStation, want games to play.[8]

Developers of games want to write for consoles that have many players.[9] Console manufacturers must attract both developers and users. Payment cards—credit, debit and charge cards—are yet another example.[10] They are used by consumers to make payments and by merchants to take payments. Merchants are more willing to accept cards that are more widely held by shoppers, and shoppers are more willing to carry cards that are accepted by more merchants. Industries that arrange for buyers and sellers to meet each other also are two-sided markets: Internet-based business-to-business exchanges (B2B), apartment rentals, and corporate bond trading are but a few examples. Dating clubs, computer-operating system makers, video-game manufacturers, payment-card systems, and B2B exchanges produce "platforms" that make "matches" between two or more distinct groups of consumers.

---

[6] A positive externality is something that party A generates for party B but for which party A has no practical way to demand compensation. Planting flowers in your front yard is an example. A negative externality like automobile exhaust fumes causes harm. *See* Paul Samuelson, *The Pure Theory of Public Expenditure,* 36 REV. ECON. STAT. 387-89 (1954). DENNIS W. CARLTON & JEFFREY M. PERLOFF, MODERN INDUSTRIAL ORGANIZATION 82-83 (3d ed. 2000).

[7] For discussion of the empirical importance of the new economy, see David S. Evans & Richard Schmalensee, *Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries,* in INNOVATION POLICY AND THE ECONOMY VOL. 2 (Adam B. Jaffe *et al.* eds., 2002).

[8] See, e.g., Chris Morris, *Console Wars: Round Two,* CNN.COM (visited Aug. 7, 2002) <http://money.cnn.com/2002/05/22/technology/e3_consolewards>. Morris states: "Console hardware sales help establish a user base, but it's ultimately the games that make or break a system."

[9] *See*, e.g., Jim Davis, *Microsoft Woos Game Developers,* CNET.COM (visited Aug. 5, 2002) <http://news.com.com/2100-1040-237819.html>. In connection with the launch of Microsoft's video game console, Davis states: "Microsoft will have to convince people (i.e., developers) that there will be an installed base that makes development efforts worthwhile (…). In other words, there have to be enough potential customers for people to sell games at a profit."

[10] DAVID S. EVANS & RICHARD SCHMALENSEE, PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING 150 (1999).

A key aspect of the business model for most of these industries involves the optimal pricing structure: the division of revenues between the two sides of the market that gets both sides on board.[11] Most computer operating system vendors do not seek significant revenues from software developers, choosing instead to collect mainly from users (Windows) or from the sale of complementary hardware (such as Palm and Sun). Sellers of video game consoles, however, do earn significant revenues from the game developers. Charge-card companies like American Express earn a disproportionate share of their revenue from merchants. And to take an example not discussed above, media sites tend to give readers content for free and collect their revenue from advertisers. The need for a pricing *structure* as well as a pricing *level* distinguishes industries based on a two-sided market from the industries ordinarily studied by economists. In two-sided markets, the product may not exist at all if the business does not get the pricing structure right.[12]

It turns out that most, if not all, industries characterized by network effects—a subject of considerable economic theorizing since the mid 1980s—are two-sided markets.[13] A network effect arises when the value that one user receives from a product increases with the number of other users of that product. A modern, but already almost quaint example, is the fax machine. You value a fax machine if there are many people to whom you can send faxes and who can send faxes to you. Most network effects arise because the product tends to be two sided. That is clear when there are two distinct types of customers (such as men and women in the dating club example). In other cases the two-sided nature it is subtler. Consider the fax machine. Most people want to both send and receive faxes. There is just one kind of customer who wants to be on a network with other customers. The same is true for telephones, instant messaging, and e-mail services. However, at any given instant, these markets can be divided into senders and receivers and the pricing strategies adopted by businesses in these markets are driven in part by these two kinds of customers.[14]

---

[11] We focus on division of the revenue since many of the costs of providing two-sided markets are joint. For documentation of the assertions in this paragraph, see the discussion of observed business models in Section II and the discussion of the theory of two-sided market pricing strategies in Section III.

[12] The district court in NaBanco noted that at an interchange fee (which, as discussed later, is used by the card systems to balance demands on the cardholder and merchant sides) of zero, the Visa system would have likely been substantially less successful. *See* National Bancard Corp. (NaBanco) v. VISA U.S.A., 596 F. Supp. 1231, 1262 (S.D. Fla. 1984).

[13] For a discussion of network effects, see generally Michael L. Katz & Carl Shapiro, *Network Externalities*, *Competition, and Compatibility,* 75 Am. Econ. Rev. 424 (1985); Michael L. Katz & Carl Shapiro, *Technology Adoption in the Presence of Network Externalities*, 94 J. Pol. Econ. 822 (1986); Chien-fu Chou & Oz Shy, *Network Effects without Network Externalities*, 8 Int'l J. Indus. ORG. 259 (1990); S.J. Leibowitz & S.E. Margolis, *Network Externality: An Uncommon Tragedy*, 8(2) J. Econ. Persp. 133 (1994).

[14] *See* Rochet & Tirole (2001), *supra* note 4, at 36 n.26. For economic models that treat the two-sided nature of communication markets, see Jean-Jacques Laffont *et al., Internet Interconnection and the Off-Net-Cost Pricing Principle,* Rand J. Econ. (forthcoming).

Industries with network effects have been under close examination by U.S. and foreign antitrust enforcement agencies.[15] Several significant cases were brought by these agencies during the 1990s and early 2000s and businesses that competed in two-sided markets were at the center of most of these cases: the AOL-Time Warner merger (U.S. and European authorities investigated two-sided markets including Internet portals, magazines, and free television);[16] the credit-card association investigations (Australian and European authorities investigated a two-sided market involving merchants and card users);[17] U.S., European and private antitrust cases against Intel (which competes in a two-sided hardware platform market);[18] the Microsoft cases (U.S. and European authorities investigated two-sided markets involving operating systems and other possible computer platforms);[19] and probes into

---

[15] For instance, in his speech Robert Pitofsky, a former FTC Commissioner, noted, "The importance of these industries [high-tech industries] to the economy, and limited antitrust experience and precedent with regard to some uses of intellectual property, suggests that there ought to be careful antitrust attention to be certain that critical economic growth is not compromised by the abuse of private market power." See Robert Pitofsky Speech before the American Bar Association, *Antitrust Analysis in High-Tech Industries: A 19th Century Discipline Addresses 21st Century Problems,* Feb. 25-26, 1999, (visited Aug. 16, 2002) <http://www.ftc.gov/speeches/pit1.htm>. Joel Klein, a former Assistant Attorney General of Antitrust Division of the Department of Justice, noted, "Civil non-merger enforcement has become especially important in this era of rapid technological change and the growth of the network industries, and we have also been very active in this area to ensure that antitrust enforcement keeps up with these changes…" *See* Joel Klein Statement before the Committee of the Judiciary of the United States House of Representatives, Washington D.C., Apr. 12, 2000 (visited Aug. 16, 2002) <http://www.usdoj.gov/atr/public/testimony/4536.pdf>. The EC Commissioner for Competition Policy Mario Monti also noted that "competition rules are all the more necessary in the era of the Internet. The Internet is a wonderful enabling technology, which will in principle, increase competition in many markets. Nevertheless, that does not mean that it is immune from competition problems." See Mario Monti Speech, *Competition and Information Technologies*, Sept. 18, 2000 (visited Aug. 16, 2002) <http://europa.eu.int/rapid/start/cgi/guesten.ksh?p_action.gettxt=gt&doc=SPEECH/00/315|0|RAPID&lg=EN>. For a discussion of four U.S. antitrust cases during the Clinton Administration that involved network industries, see David S. Evans, *All the Facts That Fit: Square Pegs and Round Holes in U.S. v. Microsoft*, 22 REG. 54, (1999).

[16] *See* Complaint, FTC Docket #C-3989 (downloaded Jul. 25, 2002) <http://www.ftc.gov/os/2000/12/aolcomplaint.pdf>. Other relevant documents may be found at <http://www.ftc.gov/os/caselist/c3989.htm> (visited Jul. 25, 2002).

[17] *Commission Plans to Clear Certain Visa Provisions, Challenge Others*, European Commission Press Release IP/00/1164, Oct. 16, 2000 (visited Aug. 14, 2002) <http://europa.eu.int/comm/competition/antitrust/cases/29373/studies/>; *Designation of Credit Card Schemes in Australia*, Reserve Bank of Australia Media Release, Apr. 12, 2001 (visited Aug. 20, 2002) <http://www.rba.gov.au/MediaReleases/2001/mr_01_09.html>

[18] *See* Matt Loney, *EC to Drop Intel Antitrust Investigation*, ZDNET, Feb. 4, 2002 (visited Aug. 20, 2002); <http://news.zdnet.co.uk/story/0,,t269-s2103680,00.html James Niccolai>; James Niccolai, *FTC Closes Intel Investigation*, PC WORLD, Sept. 26, 2000; Michael Kanellos, *Court Lifts Injunction in Intel-Integraph Case,* CNET.COM, Nov. 5, 1999 (visited Aug. 20, 2002) <http://news.com.com/2100-1040-232538.html?legacy=cnet>.

[19] *See* United States v. Microsoft Corp., Civil Action No. 98-1232(CKK), Stipulation, Nov. 2, 2001 (downloaded Jul. 25, 2002) <http://www.usdoj.gov/atr/cases/f9400/9462.htm>; and *Commission Initiates Additional Proceedings against Microsoft*, European Commission Press Release IP/01/1232, Aug. 30, 2001 (visited Aug. 20, 2001) <http://europa.eu.int/comm/competition/index_en.html>. Other relevant documents may be found at <http://www.usdoj.gov/atr/cases/ms_index.htm> (visited Jul. 25, 2002).

online broker-dealers (six separate U.S. regulatory investigations and one European investigation looking into anticompetitive behavior in two-sided e-dealer markets).[20]

In some cases, the two-sided nature of the market was central to the allegations in the antitrust case. For example, the credit-card investigations involved the pricing structure used to balance the two-sided demand[21] and the U.S. Microsoft case included the claim that one side of the market (applications) was the source of a barrier to entry.[22] In other cases, the two-sided nature of the market provided an important backdrop for understanding the workings of the business. For example, current investigations into online bond and currency exchanges are examining how dealers encourage the use of their trading platforms among buyers and sellers.[23] As another example, the European Commission was concerned that the AOL Time Warner merger would create a dominant platform in a two-sided market. The concern was that the merged company could use its allegedly dominant position in on-line music content: AOL, through its contractual agreements with Bertelsmann, a German media group, and Time Warner would have had a combined share of 30-40 percent of music content in Europe according to the Commission.[24] Before approving the merger, the Commission required the companies to take steps to eliminate AOL's contractual links to Bertelsmann, so that the Commission was satisfied that the merged firm "would not have the critical mass in terms of music publishing rights to dominate the market."[25] In the late 1980s and early 1990s, Nintendo's videogame business was also the subject of FTC and state investigations that looked at Nintendo's licensing practices and the technology it used to restrict unauthorized access to the platform.[26]

---

[20] *See* Chris Sanders, *Update 1 – BrokerTec Says Profitable since Q4 of 2001*, Reuters News, Jun. 6, 2002, in forbes.com, (visited Jun. 10, 2002) <http://www.forbes.com/newswire/2002/06/10/rtr627112.html>. The article names several online trading companies as being investigated, including the bond trader BrokerTec and currency trader FXall. *See also* Chris Sanders, *BrokerTec Confirms Probe by US Antitrust Official*, Reuters News, May 16, 2002; and *Online Trading Draws Greater Scrutiny*, Reuters News, in Cnet.com, May 17, 2002 (visited Aug. 21, 2001) <http://news.com.com/2100-1017-916334.html?legacy=cnet&tag=lh>.

[21] *See* Christian Ahlborn *et al., The Problem of Interchange Fee Analysis: Case Without a Cause?* 22(8) Eur. Competition L. Rev. 304 (2001).

[22] *See* United States v. Microsoft Corp., 253 F.3d 34, 52 (D.C. Cir. 2001). For other relevant documents see DoJ website (visited Aug. 30, 2002) <http://www.usdoj.gov/atr/cases/ms_index.htm>.

[23] *See* Chris Sanders, *Analysis—Investigators Sniff Out Online Trading, Again*, Reuters News, May 16, 2002; and Sanders, *supra* note 20.

[24] *Commission Opens Full Investigation into AOL/Time Warner Merger,* European Commission Press Release IP/00/634, Oct. 19, 2000 (visited Aug. 20, 2002) <http://europa.eu.int/comm/competition/index_en.html>. See also EEC Regulation No. 4064/89, Merger Procedure, Art. 8(2) ¶ 46 (Nov. 10, 2000).

[25] *See* EEC Regulation No. 4064/89, Merger Procedure, Art. 8(2) (Nov. 10, 2000); *Commission Gives Conditional Approval to AOL/Time Warner Merger*, European Commission Press Release IP/00/1145, Oct. 11, 2000 (visited Aug. 30, <http://europa.eu.int/rapid/start/cgi/guesten.ksh?p_action.gettxt=gt&doc=IP/00/1145|0|RAPID&lg=EN>.

[26] *See* D. Sheff, Game Over: How Nintendo Zapped An American Industry, Captured Your Dollars, and Enslaved Your Children ch. 12 (1993).

Despite their economic importance, however, two-sided markets have only recently received attention from economists and, with the exception of some recent work on payment cards, have received virtually no attention in the scholarly literature on antitrust.[27] This chapter explains the economics of two-sided markets, considers several important examples, and discusses the implications of this kind of market structure for antitrust analysis.

The remainder of this chapter is divided into four parts. Part II describes the features of two-sided market industries and explores the business models that firms servicing them tend to adopt. It finds that those firms search for a pricing structure that will solve the well-known chicken-and-egg problem common to firms in two-sided markets. It also shows that firms usually invest resources in developing both sides—sometimes by subsidizing supply on one side of the market and sometimes by producing that supply themselves.[28] Part III then delves into the economic reasons behind these business practices. It shows that businesses serving two-sided markets have to set pricing *structures* that balance the demand on the two sides of the market. The combined price charged to those on both sides of the market will have a relationship to the matchmaking firm's marginal costs of making a match. However, the prices charged on a particular side of a market will ordinarily not have a relationship either to the marginal cost of making the match or to costs specific to that side. Indeed, a key feature of these markets is that, because the product jointly benefits two parties, there is no basis for separating benefits or costs.[29] It also explains why investment in one side of the market or other, possibly through self-production, is necessary. Part IV discusses the implications of these features of two-sided markets for antitrust analysis. It shows how standard market definition, predatory pricing, vertical restraints, and coordinated effects analyses must be modified to take into account the two-sided nature of these markets. Part V presents brief conclusions.

II. SURVEY OF TWO-SIDED MARKETS

Credit cards, computer operating systems, video games, corporate bond trading, and residential real estate comprise an extraordinarily diverse set of industries. They serve different types of consumers—from adolescent boys to large retailers. Their technologies and business arrangements are quite dissimilar—compare Century 21, Microsoft, and Visa. Yet firms in these industries have adopted similar business models and pricing strategies for solving the problem they have in common—getting and keeping two sides of a market on board. We can see this from considering several different types of two-sided market industries.

---

[27] *See* Schmalensee (2002), supra note 4; and Rochet & Tirole (1999), *supra* note 4.

[28] For a related discussion, see Annabelle Gawer & Michael A. Cusumano, Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation (2002).

[29] It is well known in economics that the allocation of joint costs is arbitrary. See David M. Kreps, A Course In Microeconomic Theory 255 n.i (1990). See also Carlton & Perloff, supra note 6, at app. 3B.

Part A examines match-making services similar to the dating club. The residential real-estate industry operates much like a computerized dating service while investment bankers are more like the village matchmakers. Part B considers advertising-supported media. Here the intermediaries match groups of sellers (advertisers) with groups of buyers (consumers). Part C looks at computer operating systems. Here the intermediary provides a platform—the software—in which those who write application software and those who use application software can "meet." Part D examines video games—a market that is similar to computer operating systems but which has developed a somewhat different business model. Part E considers payment-card systems. The intermediaries—American Express for example—have something in common with the individual match makers (real estate) and the group match makers (media). Individual matches are important—I can use my American Express card to pay for something at a merchant that takes the American Express card. Group matches are important too. American Express delivers a particular class of merchants to a particular class of consumers—and vice versa.

In all these cases we shall see that the intermediary—like the dating club—helps customers complete a "transaction" by providing a "platform." The transaction occurs when members of both sides get together. Looked at from one side, a transaction is buying a house, watching a television show, buying a software application that runs on a computer, buying a video game that runs on a game console, and using a credit card at a store. The platform is the product (or set of institutional arrangements) upon which this transaction occurs. A free television show is a platform where advertisers meet consumers who may be receptive to their advertisements. A video game console is a place where people who make games can transact with people who play games. A payment-card system is a place where merchants who sell goods can meet customers who want to buy those goods.

The intermediaries succeed in the businesses discussed below by figuring out a pricing structure that internalizes the externalities between the two sides. The residential real estate industry does this by charging potential buyers nothing to look, even though each buyer imposes a cost on the system; and by charging sellers only if a sale results, even though costs are incurred when a sale does not take place. This pricing structure tends to increase the number of potential buyers and sellers just like the dating club pricing structure tends to increase the number of men and women looking for dates. Software operating system vendors internalize these externalities by reaping most of their revenue from end-users, while video game console manufacturers earn significant revenues from royalties from software game developers.

A. Market Makers—Matching Buyers and Sellers

Real estate agents, apartment finders, merger and acquisition bankers, corporate bond markets, and Internet-based business-to-business exchanges (B2B) match buyers and sellers. Each of these market makers must first solve a chicken-and-egg problem: how to attract buyers without a lineup of established sellers and how to obtain the lineup of sellers without first demonstrating a group of willing buyers. Market makers sometimes solve this coordination problem by collecting a disproportionate share of revenues from one side of the market, perhaps even taking a loss on one side of the market.[30]

Residential real estate sales in the United States is an example of a market maker that typically charges one side (the seller) while providing free (or low cost) services for the other (the buyer). Most residential property sales result from the services of a real estate agent or broker.[31] Real estate agents place the property on a common database called a multiple listing service (MLS),[32] where sellers are able to show their homes to a large audience of buyers, thus potentially increasing the number of offers that they would otherwise receive without the help of a broker.[33] Full service agents typically list the property on a local MLS as part of the overall service package without explicitly charging for the listing.[34] Real estate agents then charge the property seller a commission when the property is actually sold.[35] The traditional sales commission in the United States is six

---

[30] Because sometimes both the costs and the product are joint, the concept of "loss" on "one" side of the market is not precise. Here I refer to loss loosely as situations in which one side of the market generates a de minimis (and sometimes zero) share of the revenues and has substantial investments made in it. This will suffice for our purposes.

[31] Approximately 92 percent of the sales analyzed in studies reviewed by Zorn and Larsen were accomplished through brokers. See Thomas S. Zorn & James E. Larsen, *The Incentive Effects of Flat-Fee and Percentage Commissions for Real Estate Brokers*, 14(1) Areuea J. 24, 27 n.3 (1986).

[32] Multiple listing services are offered through local non-profit real estate boards or associations. Agents pay a one-time membership plus a monthly flat fee in order to list properties on an MLS. MLS fees are not tied to the number of properties listed and can be viewed as a fixed cost for agents. The only condition dictated by MLS membership is that agents representing sellers split the commission with agents representing buyers—the specific terms of commission-sharing arrangements, as well as the commission itself, are determined by the individual agents. See Metro MLS website (visited Aug. 30, 2002) <http://www.metromls.com/>; see also Regional MLS of Minnesota website (visited Aug. 30, 2002) <http://www.rmls-mn.com/rules.html>.

[33] *See* Metro MLS website (visited Aug. 30, 2002) <http://www.metromls.com/>. Property owners can always attempt to sell their own homes. A seller will only list with a broker if the broker has lower transaction costs. That is, if the broker can show the home to more potential buyers in a shorter span of time at a lower cost than the seller. James R. Frew & G. Donald Jud, *Who Pays the Real Estate Broker's Commission?* in Researchin Law and Economics: The Economics of Urban Property Rights 177, 178 (Austin J. Jaffe ed. 1987).

[34] However, discount real estate brokers (who simply facilitate the transaction without providing advisory services and can have commissions as low as one percent of the selling price) may add an extra 3 percent of the selling price to their fee for listing the property on the MLS. *See* Robert Erwin, *Are Discount Brokers Too Good to be True?* Wall St. J. Online (visited Aug. 22, 2002) <http://homes.wsj.com/columnists/qa/20010914-irwin.html>. The typical (non-discount) broker or agent's fee will include such services as advising the seller on a reasonable listing price and showing the property to prospective buyers, without charging extra for the MLS listing. *Id*. at 177.

[35] The individual agents set commission percentages. While all licensing is conducted through local non-profit real estate boards or associations, implying that each licensed real estate agent is a member of the local board, these boards do not dictate pricing terms to their members. *See* Regional MLS of Minnesota website (visited Aug. 30, 2002) <http://www.rmls-mn.com/rules.html>.

percent of the property's selling price.[36] In contrast, the buyer has access to all of the properties listed and can usually buy a home through an agent without paying that agent.[37] Subsidies of this sort get participation from a key side of the market—in this case, the buyer.

Residential real estate agencies have an apparently simple method for pricing—actual sellers pay a fixed percentage of the sales price. But there is a more complex method in the background. Potential buyers and sellers are not charged for access to the MLS. They also are not charged for house showings. Potential sellers pay a commission only if a sale is consummated. The prices to potential buyers and sellers do not bear any obvious relationship to the any costs that are specific to serving each side.

Three points are worth noting. First, potential buyers and sellers of real estate benefit from having someone organize a "bazaar" for them. They all have greater match prospects. Second, potential buyers and sellers have no practical way to internalize externalities from this bazaar. Each seller has benefited from all the buyers who came through his house; and each buyer has benefited from all the houses he has seen. Third, the market intermediary (the real estate agencies and the local associations that provide the MLS) appears to internalize these externalities by adjusting prices between the potential buyer side and the potential seller side. They do not charge buyers or sellers any search fees; and they impose the transaction fee entirely on the seller. Whether they have found the best possible price structure from the standpoint of operating the residential real estate industry efficiently is beyond our purview. Indeed, as we consider different industries, what is interesting is that searching for the "right" price structure is a very important aspect of building a successful business.[38]

A similar business model applies to an otherwise very different business. Companies that want to be acquired (sellers) usually retain an investment banker to seek out potential acquirers (buyers). The services of an investment banker with good contacts in key industry segments is important for making matches, especially among firms that are privately held.[39] If the firm is purchased, the

---

[36] Motoko Rich, *Sellers Trade Brokers for Internet Discounters*, Wall St. J. Online (visited Aug. 2, 2002) <http://homes.wsj.com/buysell/agentsandbrokers/20001221-rich.html>; Patrick Barta, Real Estate Agents Face a *Threat from the Web,* WALL ST. J. ONLINE (visited Aug. 2, 2002) <http://homes.wsj.com/buysell/agentsandbrokers/20020108-barta.html>; Danielle Reed, Buyers and Sellers *Decide to Go It Alone*, Wall St. J. Online (visited Aug. 2, 2002) <http://homes.wsj.com/buysell/agentsandbrokers/20010330-reed.html>.

[37] That is not true, for example, in France where residential real estate is generally not centrally listed. In the United States, buyers may end up not purchasing a house through a broker, in which case the broker does not even get paid by the seller.

[38] Apartment rental agents and location services operate in a somewhat different fashion: typically the rental property owner pays fees to the rental agent in exchange for the agent channeling renters to the property. However, in tight housing markets, such as those in New York City or Boston, the renters are required to pay the real estate broker's commission instead of the landlord. *See* Courtney Ronan, *Apartment Locaters: How Do They Make Their Money*? Realty Times website, Jun. 30, 1998 (visited Aug. 22, 2002) <http://realtytimes.com/rtnews/rtcpages/19980630_aptlocator.htm>.

[39] *See* Gabor Garai, *How to Use Intermediaries to Sell Your Business*, in Handbook Of Business Strategy 285, 288 (Pamela Goett ed., 2001).

banker is paid a transaction fee (similar to the commission fee above). Most often, the transaction fee is a portion of the purchase price,[40] although the investment banker may also set a minimum charge.[41] The seller also typically pays a retainer (an up-front fee) to the banker to cover the cost of researching potential matches based on the seller's demands.[42] Potential buyers pay nothing.[43] Similar fee structures prevail for other investment banking services, such as leveraged buyouts and private placements.[44]

Some matchmaker markets have not adopted pricing structures that are clearly biased towards buyers or sellers. Consider electronic business-to-business (B2B) exchanges, which bring buyers and sellers together usually over the Internet or through a privately operated network.[45] B2Bs usually follow either a transaction-fee-only model or a model that includes any combination of registration fees, transaction fees, and listing (or hosting) fees.[46] Registration fees may be charged to buyers, sellers, or both, and typically involve either a one-time payment or annual fees in exchange for access to the products or services of the B2B[47]—including

---

[40] One common formula for determining the broker fee in a large transaction is the Lehman formula, or 5-4-3-2-1 formula. Under this arrangement, 5 percent is paid on the first $1 million of the sale price, 4 percent on the next million, 3 percent on the next million, 2 percent on the next million, and 1 percent on the remaining excess. However, in most small transactions as well as many large transactions, a fixed percentage, equal to approximately 5 to 10 percent of the entire purchase, is generally used instead. *See* John W. Herz at al., *Broker and Finder Agreements*, in The Mergers and Acquisitions Handbook, 135, 137-8 (Milton L. Rock et al. eds., 2d ed. 1994).

[41] For example, the investment banker may set a minimum transaction fee of $500,000 for a sale price ("consideration") of up to $10 million, 1.5 percent of consideration for the first $100 million over the minimum threshold and an additional 1 percent of consideration for amounts over the first $100 million. For a $125 million acquisition, then, the fees would break out as follows: $500,000 minimum + (1.5 percent × $90 million = $1,350,000) + (1.0 percent × $25 million) = $2,100,000. If the client had paid a retainer of $50,000, this amount would be applied to the transaction fee, leaving $2,050,000 due upon deal completion. For a more general discussion, including the varying fee structures in broker-arranged M&A deals, see *id*. at 135-147.

[42] Andrew J. Sherman, Mergers and Acquisitions From A To Z: Strategic and Practical Guidance For Small- and Middle-Market Buyers and Sellers 28 (1998).

[43] Pricing is rarely this simple. Depending on the relative bargaining power of the two firms, the buyer may end up paying all or some of the transaction fee. *See* Silverstein, *supra* note 40, at 140.

[44] *See* Shelley Winston, *Producer Price Indexes for Stockbrokers, Dealers, and Investment Banking Companies*, Producer Price Index Program, U.S. Bureau of Labor Statistics (visited Aug. 14, 2002) <http://www.bea.doc.gov/bea/papers/stockbrokers.pdf >, at 10.

[45] *See generally* Steven Kaplan & Mohanbir Sawhney, *E-Hubs: The New B2B Marketplaces*, Harv. Bus. Rev. 97 (May-June 2000).

[46] *Entering the 21st Century: Competition Policy in the World of B2B Electronic Marketplaces, Part 1: Overview of B2B Electronic Marketplaces*, Federal Trade Commission Staff Report (Oct. 2000) (downloaded Aug. 14, 2002) <http://www.ftc.gov/os/2000/10/index.htm#26>, at 14-16.

[47] *Id*. at 14-16.

reduced costs of searching for an audience of buyers/sellers.[48] Transaction fees are traditionally based on either the monetary value of the transaction (and can be assessed to either buyers or sellers) or on savings realized by the buyer as a result of conducting the transaction through the particular online B2B.[49] Listing or hosting fees are generally paid by the seller in exchange for permission to market products or services over the online B2B's website; the B2B operates as a "catalog" for the seller to market its products to the B2B's audience.

The B2B's are like the dating clubs. Successful exchanges need many buyers and sellers. The potential buyers and sellers cannot internalize the benefits of having many potential parties to transact with on the other side. However, it does not seem to have been necessary—at least at this point in their development—to bias the pricing structure to one side or the other. The B2Bs seem to be able to get both sides on board without much ingenuity in either pricing or investment.

Many stock-trading systems also function without offering prices that obviously benefit one side over the other. For example, most U.S. stocks trade on auction-style systems, such as the New York Stock Exchange, in which participants offer to buy and sell at posted prices.[50] As with all exchanges, this only works if there are buyers and sellers looking to trade frequently.[51] Otherwise, an interested buyer may look to trade on the exchange but find no sellers, or vice-versa, within a reasonable time-frame. Such an exchange has little "liquidity" and would have minimal appeal to many buyers or sellers. There is enough interest in buying and selling U.S. stocks to make auction-style exchanges viable.

In contrast to stocks, corporate bonds are infrequently traded and are, at least for now, traded through dealers rather than exchanges.[52] The lack of liquidity is due to a lack of trading activity generally rather than any overall imbalance between buyers and sellers. The corporate bond dealer helps address this liquidity problem by holding an inventory of bonds. It is willing to buy a bond from a customer, and hold it in inventory to be sold later, even if a buyer cannot be located right away.[53] The dealer attempts to make money by selling at a higher price than it bought. A dealer also provides liquidity by actively trying to find buyers and sellers for different bonds.[54] In the language of two-sided

[48] *Entering the 21st Century: Competition Policy in the World of B2B Electronic Marketplaces, Part 2: Efficiencies of B2B Electronic Marketplaces*, Federal Trade Commission Staff Report (Oct. 2000) (downloaded Aug. 14, 2002) <http://www.ftc.gov/os/2000/10/index.htm#26>, at 4-7. In addition, Internet technology makes it possible for buyers and sellers in an electronic B2B marketplace to obtain a great deal of information about the B2B customer base, including number of transactions, transaction volumes, the number of other products the customer viewed before making the transaction, and customer's identity, among other things. *Entering the 21st Century, Part 1, supra* note 46, at 4-5.

[49] *Entering the 21st Century, Part 1, supra* note 46, at 14.

[50] Antony Santomero & David Babbel, Financial Markets, Instruments & Institutions 440-441, 445-449 (2d ed. 2001).

[51] *Id*. at 440.

[52] *Id*. at 439.

[53] Even on auction-style exchanges, firms known as "specialists" in specific stocks perform a similar role. *Id*. at 443-444.

[54] In this context, the dealer is also acting partly as a "broker" between buyers and sellers. For more on this, see Id. at 437-440.

markets, the dealer attempts to resolve temporary imbalances between the two sides of the market by either standing in for, or investing efforts to locate, the missing side. In essence, it is willing to provide, or find, the "chicken" or the "egg," as needed, to ensure the viability of its platform. This is in contrast to other markets we have discussed where there is a need to systematically develop one side of the market. Here, the side that is missing varies from transaction to transaction. Relative prices to buyers and sellers in dealer markets are difficult to assess, in part because prices are not generally publicly reported. There is no obvious bias in favor of buyers or sellers, which is what would be expected given the lack of a need to systematically develop a particular side of the market.

By the late 1990s, a number of firms had started developing auction-style electronic exchanges for trading corporate bonds.[55] The idea was that advances in technology (many of these exchanges operated over the Internet) might allow buyers and sellers to join a virtual exchange, on which enough buy and sell orders might be posted to generate sufficient liquidity to make the exchange viable. Some participants hoped that such an exchange, if it were successful, might offer better prices than available from dealers.[56] However, while technological advances might have allowed a larger exchange (with more liquidity) than before, the basic liquidity problem remained difficult to surmount. Customers, which are primarily large institutional investors, complained that electronic exchanges without dealer participation did not solve the liquidity problem.[57] The other side that this two-sided market needed was the dealer community, which had the liquidity that the institutional investors required. Dealers, of course, had little interest in providing their liquidity to auction-style exchanges that were designed to ultimately replace them, but many corporate bond dealers did develop their own systems. Such electronic exchanges owned or run by dealers, were regarded by some as promising,[58] although it is unclear whether they would eventually replace dealer markets with auction markets, or simply complement dealer markets.

B. Advertising-Supported Media

In market-maker businesses, the intermediaries (e.g. MLSs in real estate, Lehman Brothers in corporate bonds, CNET in B2Bs help match individual buyers with individual sellers. By contrast, in media markets, the intermediaries (e.g. Time magazine, MTV, CBS, AOL) match a group of

---

55 *The 1999 Review of Electronic Transaction Systems: eCommerce in the U.S. Fixed Income Market,* The Research Department of the Bond Market Association Report (Nov. 1999) (visited Aug. 29, 2002) <http://www.bondmarkets.com/research/ecommerce_survey99.pdf>; see also *The 2001 Review of Electronic Transaction Systems: eCommerce in the U.S. Fixed Income Market*, The Research Department of the Bond Market Association Report (Dec. 2001) (visited Aug. 29, 2002) <http://www.bondmarkets.com/research/ecommerce/ETSRprt1201.pdf>.

56 *See* Chuck Epstein, *Electronic Bond Trading Systems Pass Critical Mass*, Wall St. Tech. Online, Feb. 1, 1999 (visited Aug. 29, 2002) < http://www.wallstreetandtech.com/story/WST20000619S0015>.

57 *Bonds Get Wired*, The Forrester Report (Sep. 2000), at 5.

58 *Id*. at 13.

buyers with a group of sellers. The sellers in this case are the advertisers looking for a platform to pitch their products to a receptive group of readers or viewers. The newspapers, television channels, and Web sites that form media platforms are "audience makers" rather than "market makers."[59] That is, advertising-supported media providers are interested in attracting advertisers on one side and subscribers on the other to form an audience for the advertisers.

Like many market makers, most audience makers earn a disproportionate share of their revenues from one side of the market and, like some market makers, may lose money on one side of the market. Newspapers, magazines, and television networks charge advertisers a placement fee dependent on the size or length of the ad and on the circulation or subscriber base of the media platform.[60] For example, full page ads cost more than quarter page ones and television ads that run during hit shows or special programs, like the Super Bowl, cost considerably more than those that run during non-prime-time slots. For instance, during the 2002 Super Bowl, which ran on FOX and attracted 86.8 million television viewers, the average price for a thirty-second commercial was $1.9 million.[61] In contrast, the average price for a thirty-second commercial on FOX is slightly over $150,000.[62]

The fees that media platforms collect from advertisers pay for the content that the media presents to the audience. That is, FOX buys the rights to shows such as *Malcolm in the Middle* with proceeds from advertising, and in exchange, advertisers are allotted a specific amount of time during the show to market to the show's audience (which may include a somewhat specialized demographic group).[63] Contracts between the advertiser and television media platform generally include

---

[59] "In a fundamental sense, what advertisers demand, and what the various advertising media outlets supply, are units of audience for advertising messages. Thus advertiser demand for space in the print media and time in the broadcast media is a derived demand stemming from a demand for audience, and is a positive function of the size and quality of audience." James M. Ferguson, *Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition*, 26 J.L. Econ. 635, 637 (1983).

[60] Newspapers also have different rates for different classes of advertising, such as display versus classified, and different ad types, such as black-and-white versus colored. Fees can also vary depending on the day of the week (Sunday charges may be higher) and time of day (evening edition versus morning edition). A typical newspaper charge is determined by the "milinch," the column inch per thousand circulation. *Id*. at 653. For television, approximately 70 to 80 percent of network television advertising time is sold prior to the start of the television season, through contracts specifying price and guaranteed minimum rating of the television shows. Ronald Goettler, *Advertising Rates, Audience Composition, and Competition in the Network Television Industry*, Carnegie Mellon University Gsia Working Paper #1999-E28 (Aug. 8, 1999), at 4.

[61] *TV, Internet and Advertising Trends Surrounding the 2002 Super Bowl*, Nielson Media Research (Feb. 19, 2002) (downloaded Aug. 15, 2002) <www.nielsenmedia.com/newsreleases/2002/2002%20Super%20Bowl_WriteUp-POST.doc>.

[62] *Prime-Time Pricing Survey*, Advertising Age (2002).

[63] A recent survey of by the Magazine Publishers of America (MPA) found that 45.1 percent of magazine content consisted of advertisements, while 54.9 percent comprised the editorial content. Providing advertising space generates revenue for the magazine and providing editorial content helps retain a reader base. *See* Historical Advertising /Editorial Ratios, MPA website (visited Aug. 20, 2002) <http://www.magazine.org/resources/fact_sheets/ed2_9_02.html>.

the price paid by the advertiser for commercial time and a minimum rating for the television show that is guaranteed by the media platform.[64] Advertising prices vary with the television show's audience size, age, and gender make-up.[65] For example, higher advertising prices are generally associated with a large and primarily homogenous viewing audience, along with a higher proportion of viewers between 35 to 49 years of age.[66]

On the other side, free television viewers pay only an implicit price: the cost of having to watch commercials or waiting for the show to resume. Other platforms, such as newspapers and magazines, charge the audience an explicit price. But even here readers are heavily subsidized as advertisers pay the bulk of the costs of obtaining the content that attracts their desired subscriber demographics.[67]

Audience-maker businesses have the basic characteristics of two-sided markets.[68] Their pricing structure does not have a direct relationship to the marginal cost of providing the media or its content to either side of the market. The economist's usual "marginal revenue equals marginal cost" condition does not help us understand why television viewers pay nothing for watching content that was expensive to create or why many newspapers and magazines are distributed for less than the marginal cost of production and distribution. The audience-makers are helping buyers and sellers internalize externalities. The advertisers benefit more the larger the number of potential buyers they can reach. The potential buyers realize a negative externality from advertising messages—most people would pay to avoid them; the more messages they receive the bigger the externality. These externalities seem to be addressed by a pricing structure in which advertisers pay for the audience, and the audience is paid with content for putting up with the

---

64 Television show ratings are measured by Nielson Media Research. Goettler, *supra* note 60, at 4.

65 For instance, television shows like *Roseanne* and *Melrose Place* that had desirable viewer demographics (large audience with a high proportion being female and between 18 to 49 years of age) also had substantial gains in ad revenue for each additional million viewers. This is compared with less considerable increases in ad revenue per one million additional viewers for *Hat Squad*, a show with less desirable viewer demographics (audience with a smaller proportion being female compared to *Melrose Place* and a large proportion being 50 years of age and older). *Id*. at 3.

66 *Id*. at 16.

67 Approximately 80 percent of newspaper revenue comes from advertisers. Lisa George & Joel Waldfogel, *Whom Benefits Whom in Daily Newspaper Markets?* NBER WORKING PAPER #7944 (Oct. 2000), at 11.

68 The two-sided nature of advertising is well-understood in the advertising literature. "[M]edia firms, such as newspapers, magazines and commercial television channels, operate simultaneously in two sub-markets. Not only they sell their products to readers, viewers or listeners, they also sell advertising space to firms. Moreover, these markets are generally interrelated on the demand side. For example, the value of placing an ad in a local newspaper depends on the paper's circulation, and the subscribers' valuation of the newspaper is, at least to some extent, affected by the type and amount of advertising." *See* Jonas Hackner & Sten Nyberg, *Price Competition, Advertising and Media Market Concentration*, STOCKHOLM UNIVERSITY RESEARCH PAPERS IN ECONOMICS #2000:3 (2000), at 1. *See* also KYLE BAGWELL, ED., THE ECONOMICS OF ADVERTISING Pt. VII-VIII (2001). This paper shows that advertising is part of a much broader class of two-sided markets. Also, the theoretical results discussed below do not appear to have been developed in the literature on the economics of advertising.

ads. Again, whether this is the most efficient method is beyond the scope here, although its long-term survival in many media suggests that it is. An interesting issue is whether this business model will work for Web portals. Advertisers seem reluctant to pay much for the occasional click while browsers seem reluctant to pay for content.[69]

C. Computer Software

Computer operating systems are two-sided markets comprising applications software developers and people who use the operating system—usually in conjunction with one or more software applications programs. A typical applications program consists of lines of interrelated code that carry out various tasks necessary for accomplishing whatever purpose the software was designed to do—word processing, game playing, or statistical analysis.[70] Many of the tasks that different kinds of applications must accomplish are similar—drawing dialog boxes on the screen, saving documents, and providing "help" information to users. Therefore, it is possible for the computer industry to reduce the duplication of effort across software developers by having common tasks performed by the operating system rather than each program. This is accomplished by having code in the operating system that accomplishes these tasks and interfaces that enable the software developers to use this code.[71]

Of course, the applications software developer can rely on the operating system code only if the user has this operating system running on her computer. She is more likely to have this operating system if many applications are available that she wants to use that rely on this operating system. To be successful, then, the operating system vendor has to persuade many applications developers to rely on its operating system in writing software and many software users to install this operating system on their computers. All computer operating systems— from the Palm OS used in handheld devices to Windows and Linux on personal computers to Solaris on servers—must appeal to users and developers.

---

[69] "[M]any traditional advertisers have yet to be convinced of the cost effectiveness of spending their marketing budget on the web." See *Can Dot.Coms Make Money from Advertising?* BBC News, Mar. 8, 2001 (visited Aug. 27, 2002) <http://news.bbc.co.uk/1/hi/business/1208861.stm>. Many Internet media companies have already been looking for other sources of income apart from online advertising. For instance, in 1997 Microsoft's online magazine, Slate, planned to begin asking its readers to pay $19.95 a year for the magazine they used to receive for free. Eventually, Slate abandoned this idea because it could lock out too many potential readers. See *Slate Clean of Subscription Fees*, CNET.com, Jan. 10, 1997 (visited Aug. 27, 2002) <http://news.com.com/2100-1023-261158.html?tag=bplst>. In general, consumers are reluctant to pay for most information on the Web including Webzines and music. See *Pay for Music Online? No Way!* ZDNet May 17, 2000 (visited Aug. 27, 2002) <http://comment.zdnet.co.uk/story/0,,t479-s2113872-p2,00.html>.

[70] For definitions, see NetLingo The Internet Dictionary (visited Aug. 21, 2002) <http://www.netlingo.com/inframes.cfm>.

[71] *Id.*

There is nothing preventing operating system sellers from collecting revenues from both sides of the market. Users can be charged license fees for the software.[72] Software developers can be charged for information, tools, and other services necessary for accessing the code they want to rely on in their programs.[73] In fact, all operating system vendors have chosen to get most of their revenues from the user-side of the market; vendors differ in the extent to which they invest in the applications-side of the market.[74]

Consider Microsoft. It licenses its Windows operating system to computer manufacturers (and also sells some copies at retail and through other channels). This is the source of most of the considerable revenue that it realizes from Windows.[75] Microsoft does not charge software developers for information on using the features of the operating system. Developers, in fact, have gotten numerous development tools and considerable support for free from Microsoft, including the free Microsoft Developer Network (MSDN) Web site and free software development kits (SDKs) for Windows.[76] The company also shares early beta versions of new operating system releases with developers.[77] The company holds annual development conferences to demonstrate how the Windows programming interfaces are evolving.[78] It does earn some revenues, however, from developers. It offers five different development packages ranging from $99 to

---

[72] For instance, Microsoft Corp. charges end-users a $199 fee for Microsoft Windows XP Home Edition. *See* Microsoft Corp. website (visited Aug. 21, 2002) <http://www.microsoft.com/windowsxp/order/>. Apple's charge for Mac OS 10 X v10.2 is $129. See Apple website (visited Aug. 21, 2002) <http://www.apple.com/macosx/>.

[73] For example, Xtras.net supplies certain development tools for .NET at discounted prices. *See* Xtras website (visited Aug. 21, 2002) <http://www.xtras.net/>. Also, companies that make development tools, such as Oracle, Borland, IBM, and Rational Software offer to rent web-based tools to developers. See Wylie Wong, *For Rent: Software Development Tools*, Cnet.com, Jan. 30, 2001 (visited Aug. 21, 2002) <http://news.com.com/2100-1001-251751.html?legacy=cnet>.

[74] For general commentary on the importance of investment in third party software developers, see FM Interview with Linus Torvalds, *What motivates free software developers?* FIRST MONDAY, Mar. 2, 1998 (visited Jul. 31, 2002) <http://www.firstmonday.dk/issues/issue3_3/torvalds/index.html>; Jerry Pournelle, *Jerry's Take on the Microsoft Decision: Wrong!* BYTE.COM, Nov. 8, 1999, (visited Aug. 9, 2002) <http://www.byte.com/documents/s=200/byt19991108s0001/index.htm>; and Richard Shim & Sandeep Junnarkar, *Palm Reaches Out to Developers*, CNET.COM, Oct. 23, 2001 (visited Aug. 13, 2002) <http://news.com.com/2100-1040-274797.html?legacy=cnet>.

[75] *See* Microsoft Corp. 10-K (FY2001) (downloaded Aug. 30, 2002) <http://www.sec.gov/Archives/edgar/data/789019/000103221001501099/0001032210-01-501099.txt>.

[76] MSDN website provides online support for developers including technical programming information (sample code, documentation, technical articles, and reference guides), troubleshooting, access to Microsoft MSDN magazine and other information. *See* Microsoft Corp. website (visited August 30, 2002) http://msdn.microsoft.com/ >.

[77] *See* GAWER & CUSUMANO, *supra* note 28, at 150-51.

[78] For example, the Windows .NET Server Developer Conference in Seattle, Washington from September 3–6, 2002 was focused on providing various technical content to architects, enterprise developers, ISV developers, and technical decision makers. The cost to register for the entire conference was $1,295. *See* Windows .NET Server DevCon, Microsoft Corp. website (visited Aug. 22, 2002) <http://microsoft.com/misc/external/serverdevcon/>.

$2,799 per individual license.[79] The packages include varying levels of access to and discounts on technical resources and support, code samples, documentation, development tools, software, and hardware.[80]

Microsoft has also encouraged end-user demand by developing applications programs that run on its operating system. In many cases, Microsoft has done this for the same reason any applications developer does this—to make money from the sales of a successful application.[81] In other cases, Microsoft has produced applications to encourage users to license its operating system (which in turn encourages other applications developers to write applications). In fact, during the early years of Windows Microsoft had a difficult time persuading other developers to write applications for it. In 1992 there were only 1,438 applications available specifically for Windows compared with 22,328 for DOS and 4,213 for Apple's MacOS.[82] As Bill Gates puts it, "In 1989, I personally went to all the applications developers and asked them to write applications for Microsoft Windows. They wouldn't do it. So I went to the Microsoft Applications Group, and they didn't have that option."[83]

Apple, which makes both the Macintosh computer and its operating system MacOS, has followed a similar approach.[84] Apple provides interfaces (the APIs, or application program interfaces) for its operating system to developers at no charge.[85] They also provide free SDKs that developers can download from their website.[86] However, Apple, like Microsoft, does impose some charges for tools that help software developers use the APIs. The Apple Developer Connection (ADC) program provides software and hardware developers access to development tools, software, kits, and reference materials, with membership fees ranging from the $99 per year Student Program to the $3,500 per year Premier Program.[87] Based on various membership levels, in addition to development materials and support, members of the ADC may also have access to discounted Apple hardware, discounted technical and business services, and Apple's Worldwide

---

79 Discounts are offered on volume licenses and academic licenses orders. *See* MSDN Subscriptions Pricing, Microsoft Corp. website (visited Aug. 13, 2002) <http://msdn.microsoft.com/subscriptions/prodinfo/pricing.asp#univ>.

80 *See* MSDN Subscriptions Level & Feature Overview, Microsoft Corp. website (visited Aug. 13, 2002) <http://msdn.microsoft.com/subscriptions/prodinfo/levels.asp>.

81 For example, the current retail price for the 2002 version of Microsoft Word is $339 (the upgrade from an older version costs $79.95). *See* Look Up Prices, Microsoft Corp. website (visited Aug. 21, 2002) <http://www.microsoft.com/office/howtobuy/pricing.asp>.

82 Software Product Specifications from Computer Select Cd (Jan. 1992).

83 See Pournelle, supra note 74.

84 This and the following sentences in this paragraph are based on information available in the developer section of Apple's website (visited Jul. 31, 2002) <http://developer.apple.com/membership/descriptions.html>.

85 *See* Development Kit, Apple website (visited Aug. 21, 2002) <http://developer.apple.com/sdk/>.

86 *Id.*

87 *See* Apple Developer Connection Programs, Apple website (visited Aug. 21, 2002) <http://developer.apple.com/membership/descriptions.html>.

Developers Conference. Like Microsoft, Apple earns relatively little of its revenues from the developer side of the business despite the large costs of writing operating system software code that these developers use.[88] Apple, however, has been less assiduous at courting developers than Microsoft.[89]

Palm, which is the leader in the market for handheld personal digital assistants (PDAs), currently follows a similar business model although it took a different approach to solve the chicken-and-egg problem in the beginning.[90] In 1996 Palm introduced the PalmPilot PDA, developed a new operating system to run on it, and also invested in designing application programs.[91] After obtaining critical mass on one side of the market (users), Palm reduced its investments in applications, focused on the hardware and operating system, and began to attract independent developers for applications.[92] Palm, in effect, solved the chicken-and-egg problem by initially providing its own chicken.[93] By supplying one side of the market itself, Palm could focus all of its efforts on attracting end-users. Now, however, it expends considerable effort on wooing software developers by hosting developer forums, providing technical training, and even supplying marketing for complementary software developers.[94] In September 2000, Palm launched Palm Ventures, a $50 million venture unit to aid firms developing complementary products.[95] For individual developers, Palm provides free development kits, product images, limited access to source code, and early access to tools and information.[96] Palm also offers an advanced plan for a $495 annual fee, which includes direct technical support, marketing, and quarterly updates on development tools and technology.[97]

---

[88] In 2001, Apple derived 82 percent of its total revenue from sales of Power Macintosh, PowerBook, iMac and iBook to end-users, and 18 percent from third-party developers, services, and sales of iPad, various peripheral products and software. See Apple Computer, Inc. 10-K (FY2001) (visited Aug. 27, 2002) <http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=aapl&script=800&layout=8>.

[89] *See* Pournelle, *supra* note 74.

[90] *See* Shim & Junnarkar, *supra* note 74; Media Backgrounder, Palm website (visited Aug. 13, 2002) <http://www.palm.com/about/pr/background.html>.

[91] *See* Gawer & Cusumano, *supra* note 28, at 190-91, 194.

[92] *Id.*

[93] Palm's approach was, at least in part, dictated by earlier PDA failures by Apple with the Newton and by Palm with the Zoomer. As Palm cofounder Donna Dubinsky explained, they were determined not to rely on outside developers until they had sold at least a million devices. "We are going to prove it to developers," she stated, "We are not going to ask them to trust us." *See* D.B. Yoffie & M. Kwak, Judo Strategy Ch. 5 (2001). Gawer and Cusumano argue that a do-everything strategy, while helpful initially, is not a viable long term strategy. "If the product turns out to be successful with customers, specialist firms are likely to jump into the market. Often, they become better at specific tasks (like hardware design or applications design) than firms that try to do everything." *See* Gawer & Cusumano, *supra* note 28, at 249.

[94] *See* Related Developer Programs, Palm website (visited Aug. 15, 2002) <http://www.palmos.com/dev/programs/>.

[95] *Palm Inc. Lauches VC Unit to Aid Complementary Firms*, Boston Globe, Sep. 7, 2000.

[96] Shim & Junnarkar, supra note 74.

[97] *Id.*

In addition, developers can pay to attend brief courses, such as the two-day "Advanced Palm OS Development and Debugging" course for $1,000.[98,99]

Palm's efforts appear to have paid off: in March 2002, Palm boasted that it had nearly 200,000 registered software developers and more than 13,000 software titles already available.[100] Applications range from electronic book-reading with thousands of e-books, to document-editing programs that work with Microsoft Excel and Word files, to enterprise software tools that let business users fill out forms and check inventories through a wireless connection.[101] For consumers, Palm OS powered devices typically start at $99.[102] But like Microsoft and Apple, Palm earns the preponderance of its revenues from the user-side of the business.[103]

Computer operating system characteristics are similar to those we have seen in our review of other two-sided markets. The business is about the search for the right price structure that will get both sides on board. The price structure is one that is not readily explained by the approach economists apply to one-sided markets. Here one side seems to get extraordinary benefits from the operating systems but pay little for it. Once again there is an externality problem that only an intermediary between the two sides can solve. Users benefit when they have many applications available to run on their computers and when the similar portions of those applications are common. They have no way to pay software developers for the benefits of this coordination. Likewise, developers benefit when they have a larger group of users for their software and when they can avoid duplicating the code used by other packages. They have no way to pay users for the benefits of this coordination. The intermediary can solve the problem—internalize these externalities between the two groups of customers—by adopting a

---

[98] *See* Developer Training, Palm website (visited Aug. 13, 2002) <http://www.palmos.com/dev/training/instructor/advanced.html>.

[99] As another means to encourage independent developers to write applications for Palm OS, Palm licenses its operating system to other hardware manufacturers. For example, Handspring, which makes PDAs that compete with Palm's, is one of many licensees of the Palm OS. See PalmSource website (visited Aug. 13) <http://www.palmos.com/licensees/>. Currently, both Palm and Handspring make the interface specifications for their computing platforms available at no charge to developers. Thus, developers profit from selling Palm- and Handspring-compatible software, and in exchange, Palm and Handspring profit from having a platform that becomes increasingly popular to consumers as more compatible software is developed. *See* Douglas Lichtman, *Property Rights in Emerging Platform Technologies*, XXIX J. Legal Stud. 615, 616 (2000).

[100] *See* Media Backgrounder, *supra* note 90. In recent years, Palm has faced increasing competition from the Pocket PC platform. *See* Mark Walser, *The (M)empire Strikes Back*, M Commerce Times, Jul. 9, 2001 (visited Aug. 29, 2002) <http://mcommercetimes.com/Technology/147>.

[101] See Product Overview, Palm website (visited Aug. 13, 2002) <http://www.palm.com/about/corporate/productoverview.html>.

[102] *See* Developer Training: Class Topics, Palm website (visited Aug. 13, 2002) <http://www.palmos.com/platform/os5/topten.html>.

[103] *See* Palm, Inc. 10-K (FY2002) (visited Aug. 27, 2002) <http://media.corporate-ir.net/media_files/nsd/palm/reports/fy0210k.pdf>.

pricing structure that recognizes the mutual benefits. As before, whether the pricing structure the operating system vendors have settled on is the best one for solving this problem is not readily known.

D. Video Games

While similar to operating systems in that they involve hardware and software, video game platforms have an informative history.[104] In the early days of the home video game industry, companies manufacturing game consoles did not face chicken-and-egg problems in a strict sense—consoles were essentially single-game devices, and games were hardwired into the console's circuitry. By definition, the manufacturer of the console was also the manufacturer of the game so there was no need to court independent game developers. Examples of this business model are Atari's Home Pong (1975), and Coleco's Telestar (1976).[105] This was a one-sided market.

With the release of Fairchild's Channel F in 1976 a new business model emerged. Channel F did not hardwire games, but rather played games stored in interchangeable cartridges. Atari expanded on Fairchild's approach with the release of the Video Computer System (VCS) in 1977.[106] Neither Fairchild nor Atari immediately contracted with independent game developers, but these companies did establish the separation of hardware and software in the home video game market.[107] Like Palm in the early days of the PDA industry, video games in the late 1970s solved the chicken-and-egg problem by supplying the chicken.

While still not faced with coordinating both sides of the market, the separation of hardware and software did present console manufacturers with a new business decision: how to price each component. Atari chose to sell the hardware at a relatively low price and earn a larger share of revenue from software sales. "Give away the razors so that you can sell the blades" became an axiom in the industry from this point on.[108] A large library of high-quality games was indispensable for ensuring the success of a home video game console.[109]

A new dimension of the business model started taking shape in 1980. In that year some of the Atari programmers defected and founded Activision, the first independent software company

---

[104] It should be emphasized that we focus on the history of home video games in the U.S. Some of the platform launches we describe for the U.S. were preceded by releases in Japan.

[105] *See* S. Kent, The Ultimate History of Video Games Ch.1-8 (2001). The Odyssey, manufactured by Magnavox and introduced in 1972, was the first home video game console. It was the only multi-game console of this early period. While Home Pong played only one game, Odyssey could play twelve. *See* KENT *supra*, at 80-81.

[106] The Video Computer System was also known as the Atari 2600. *Id*. at xiii.

[107] *Id*. at 94-98.

[108] *Id*. at 107.

[109] On the significance of game availability for the success of a console, see, among others, Morris, *supra* note 8. See also David Becker, *Who Will Be the King of the Game Consoles*, CNET.Com, Nov. 21, 2000 (visited Aug. 5, 2002) <http://news.com.com/2100-1040-248875.html>.

in the home video game market.[110] The company's purpose was to create games for the VCS. Activision soon released its first games and was an overnight success. Again like Palm, the VCS game platform reached such a level of success with customers on one side of the market (end-users of games) that it generated demand from customers on the other side (independent game developers). Activision represented a turning point in the video game industry: from this point on console manufacturers entering the market had to attract both gamers and developers to the platform to ensure its success. A two-sided market emerged.

Although it is unclear whether Atari or any other company in the early 1980s charged licensing fees to independent developers, from the mid-1980s onward (and more specifically from the time Nintendo entered the home video game market on), licensing fees from software developers became a major source of revenues for console manufacturers.[111] In fact, the business model that developed at least from the mid-1980s on was based on selling the console to end-users near or less than its marginal cost of production while relying on revenue from in-house-produced software and from license fees charged to independent developers to recover fixed development costs and earn a profit.[112]

Starting in the mid-1980s, Nintendo and its 8-bit Nintendo Entertainment System (NES) displaced Atari as the video game leader. Nintendo was already a big arcade company prior to entering the home segment of the market, and it attracted end-users by converting many popular arcade games into home video games.[113] As far as courting independent developers, Nintendo sought quality over quantity. Nintendo protected the NES with a security chip that locked out unauthorized, unlicensed cartridges. The licensing agreements that it signed with independent developers imposed hefty royalty rates, established that they would only publish five games per year and that those games would be exclusive to the NES for a two-year period.[114]

By the late 1980s a new competitor, Sega, emerged. Sega's strategy to compete with Nintendo was manifold. In the first place, it beat Nintendo to market with the release of a 16-bit console that

[110] Still other Atari programmers left the company to form Imagic. Like Activision, Imagic enjoyed success very quickly. *See* KENT supra note 105, at 192-193, 197.

[111] *See*, e.g., *Software Can Raise Platform Values,* Video Games & PC Entertainment Software – Marketers, The Investext Group, Jun. 1, 1998, *available in* NEXIS. *See also* Morris, *supra* note 8.

[112] *See*, for example, David Becker, *Revenue from Game Consoles will Plunge, Report Predicts*, CNET.COM, Jul. 31, 2000 (visited Aug. 5, 2002) <http://techrepublic-cnet.com.com/2100-1040-243841.html?legacy=cnet>. *See* also David Becker & Joe Wilcox, *Will Xbox Drain Microsoft?* CNET.COM, Mar. 6, 2001 (visited Aug. 5, 2002) <http://techrepublic-cnet.com.com/2100-1040-253654.html>.

[113] *See* KENT, *supra* note 105, at 278, 296-297, 299-300.

[114] *Id*. at 308-309, 351. Nintendo did grant special favors to some of its most successful licensees. It gave a second license to both Konami and Acclaim, two of the top independent game developers for the NES, so that each company would be able to produce 5 additional games per year. *Id*. at 422. On the relationships between Nintendo and its licensees, see also SHEFF, *supra* note 26, at chapter 10.

was considerably more powerful than the NES.[115] Sega's Genesis was launched in America in 1989, almost two years earlier than Nintendo's 16-bit console, the Super NES.[116] Second, in order to attract end-users Sega soon cut the price of the console from $189 to $149.[117] Third, since Nintendo dominated the arcade translation business, Sega signed contracts with athletes and celebrities in order to create a large library of games based on their names and images.[118] Fourth, Sega launched an aggressive marketing campaign to highlight the technological superiority of the Genesis over the NES.[119] Finally, even though in general the Sega licensing deals with independent game developers were similar to Nintendo's,[120] Sega cut deals with some key developers (e.g., Electronic Arts) by offering them special licensing terms.[121] Sega's strategy paid off and it outsold Nintendo in 1991.[122] Even after Nintendo launched its 16-bit Super NES, Sega remained competitive for some time on the basis of its larger library of game and its faster pace of game publishing.[123]

In 1995 yet another competitor joined the fray: Sony's 32-bit PlayStation. Sony took advantage of a window of opportunity that arose in the mid 1980s. At this time Sega was struggling in an attempt to support several incompatible platforms and Nintendo was starting to develop a 64-bit console (Nintendo 64), which it would not release until 1996.[124]

The Sony model was different from both Sega's and Nintendo's in that Sony made a much stronger effort to attract independent developers. For example, Sony offered excellent development tools to third-party publishers and PlayStation developed a reputation for being very easy to program.[125] Combined with a liberal $10 per game licensing fee and Sony's aggressive marketing

---

[115] Sega had unsuccessfully released an 8-bit console, the Master System, in 1986. *See* Kent, *supra* note 105, at 401, 404-405.

[116] *Id*. at 404, 431-434.

[117] *Id*. at 404, 427, 433.

[118] *Id*. at 404-406, 426. Sega did have its own portfolio of arcade games, but Nintendo had exclusivity on some of the most popular arcade products.

[119] *Id*. at 426-427.

[120] *Id*. at 381, 409-410. Independent developers paid Sega between $10 and $15 per cartridge on top of the real hardware manufacturing costs. Like Nintendo, Sega created security systems in its consoles to guard against unlicensed publishers. *Id*. at 381-382.

[121] In 1989 Electronic Arts' technicians successfully reverse-engineered both the Nintendo NES and the Sega Genesis. Electronic Arts rejected Sega's licensing terms and decided to go ahead and start publishing their own games on Genesis. Before proceeding, however, they offered Sega the possibility of entering into an agreement with special licensing terms. Sega accepted and granted Electronic Arts a licensing contract with increased flexibility and lower royalty rates. *Id*. at 408-410.

[122] *Id*. at 434.

[123] *Id*. at 447-448.

[124] *Id*. at ch. 27.

[125] Sega also launched its 32-bit platform, the Saturn, in 1995, but at launch time this platform had already developed a reputation for being very difficult to program. *Id*. at 509, 516. As a result of Saturn's failure, Sega stopped manufacturing consoles in early 2001. *Next Generation of Gaming: Forecasts and Analysis, 2000-2005*, IDC REPORT #24432 (Apr. 2001), at 26-28.

plans, PlayStation became a very attractive platform for game designers.[126] By the time it launched in the U.S., about 100 game developing companies had signed agreements with Sony—and more than 300 projects were either planned or underway.[127] Game developers were now writing similar titles for multiple game platforms (Nintendo, Sega, and Sony).[128] As far as the end-user side of the market is concerned, Sony attracted consumers by offering a large library of games almost from the beginning and selling the console at a very competitive price.[129]

Currently, video game business models resemble the operating system approach in that companies entering the market make significant efforts to attract independent game developers. In the process of promoting the Xbox console, for example, Microsoft announced two programs, the Independent Developer Program and the Incubator Program, to encourage smaller developers by providing free software tools and waiving normal pre-publishing requirements.[130] Furthermore, Microsoft had extensive meetings with developers before the hardware specs for the console were set and incorporated many of their suggestions into the final design.[131] Microsoft also made it easier for developers with PC experience to develop games for the Xbox by relying on DirectX (a collection of APIs that serves as the foundation for most PC games)[132] in the design of the console.

The video game history illustrates how seemingly similar two-sided markets can arrive at different pricing structures. It is possible that slight differences in technology and demand lead to different pricing structures for internalizing externalities. There is another possibility though. Coming up with the right pricing structure is a difficult problem that requires more information than setting the price of toothpaste. It may take time and experimentation for industries to converge on the optimal structure.

E. Payment Systems

Payment systems—cash, checks, cards, and emerging e-pay systems—are viable only if both buyers and sellers use it. If buyers wanted to use cash but sellers did not want to take cash—a still infrequent but not unheard of situation—then cash would not be a viable system. A payment system is

---

[126] Sony's licensing structure was based on a $10-per-game arrangement that included manufacturing disks, manuals, and packaging. Nintendo's game consoles were based on cartridges, which were much more expensive for game developers than CDs. At the time, it cost more than $20 to manufacture an 8-megabyte cartridge, but it cost less than $2 to press a 640-megabyte CD. *Id*. at 511.

[127] *Id*. at 504.

[128] *See* The Investext Group, *supra* note 111. *See* also Morris, *supra* note 8.

[129] *See* Kent, *supra* note 105, at 516-520. Sega's Saturn was launched with a $399 price tag. Sony's PlayStation was launched at $299.

[130] *Microsoft Embraces the Worldwide Independent Video Game Developer Community*, Microsoft PressPass, (visited Aug. 2001, 2002) <http://www.microsoft.com/presspass/press/2000/Nov00/XPKPR.asp>.

[131] *See* Becker, supra note 109.

[132] *See* Davis, *supra* note 9. *See also* Becker, *supra* note 109.

more valuable to sellers if more buyers take the tender and more valuable to buyers if more sellers take the tender. Governments have helped solve the chicken-and-egg problem in the case of money (of the sort that has no intrinsic value) by passing laws that require businesses and people to accept money for discharging debts.[133] More interesting from our standpoint are payment mechanisms in which businesses have had to figure out ways to get both sides of the market on board.[134]

Diners Club was the first charge card that many people used to pay at many businesses. It was introduced in 1950.[135] The card (then a slip of cardboard) was distributed for free to well-off Manhattan residents for the purpose of paying at restaurants. There was no fee. There were also no charges if the cardholder paid within a certain period of time; that meant the cardholders paid nothing for the float they received from the time of their meal to the time they paid. Using its attractive clientele as a lure, Diners Club then persuaded restaurants to accept payment with the card, charging the restaurants a cut of the tab—about 7 percent.[136] In its fourth year of operation, Diners Club added a membership fee of $34 per year.[137]

This business model has persisted for the last half century for the pure charge card (no credit, pay in full within 30 days). Consider the American Express charge card. In exchange for the rights to accept the American Express card, merchants are typically assessed a fee (known oddly enough as the "merchant discount") that is about 2.7 percent of the purchase amount on average.[138] Cardholders pay, on average, a $34 annual fee for an American Express charge and credit cards.[139] The cardholders with credit cards that have no annual fees, such as *Blue and Optima*,

---

[133] Until 1971, the gold standard system constrained the amount of money that government could issue to its gold reserves. *See* Michael D. Bordo, *Gold Standard*, The Concise Encyclopedia of Economics Online (visited Aug. 19, 2002) <http://www.econlib.org/library/Enc/GoldStandard.html>. Under this system, paper money could be exchanged for gold, which had intrinsic value. A few years before the gold standard was finally and completely abandoned, Congress passed the Coinage Act of 1965, stating that, "All coins and currencies of the United States… shall be legal-tender for all debts, public and private, public charges, taxes, duties and dues." See Legal Tender: A Definition, Bureau of Engraving and Printing (visited Aug. 16, 2002) <http://www.bep.treas.gov/document.cfm/18/110>. In essence, before officially removing the value from the currency, government passed a law requiring people to accept it for payment of debt. However, this does not mean that merchants are under legal obligation to accept cash for payment. For example, some businesses do not take pennies and certain merchants do not accept cash and only allow credit card transactions. Thus, the laws encourage the use of cash generally but did not specifically mandate merchant acceptance, leaving businesses free to form their own payment guidelines. *See* FAQs: Currency, U.S. Treasury (visited Aug. 21, 2002) <http://www.ustreas.gov/education/faq/currency/legal-tender.html>.

[134] *See generally* Evans & Schmalensee, *supra* note 10.

[135] *Id*. at 62.

[136] *Id*. at 181.

[137] A $34 membership fee is reported in 2002 dollars and is equal to $5 charged by Diners Club in 1954. *See* Peter Z. Grossman, American Express: The Unofficial History of the People Who Built The Great Financial Empire 263 (1987).

[138] *See* American Express Company Annual Report 2001 (visited Aug. 15, 2002), <http://www.onlineproxy.com/amex/2002/ar/pdf/axp_ar_2001.pdf>, at 37.

[139] *Id*.

benefit from the float.[140] Overall, excluding finance charge revenue, American Express earned 82 percent of its revenues from the merchant side of the business in 2001.[141]

The institutions and pricing mechanisms for getting both sides of the market on board are more complicated in the case of "cooperative" card systems.[142] MasterCard and Visa are associations whose members consist of banks that provide payment services to individuals ("issuing") or merchants ("acquiring") or both. In the case of individuals, the services include providing a card that can be used to make payment at merchants and that may provide some long-term credit—the issuing bank makes arrangements to pay the merchant and then bills the cardholder and may offer to finance the purchase. In the case of merchants, the services include providing technology for processing card transactions and paying the merchant. Frequently, the issuing and acquiring banks differ. The associations provide coordination: they operate networks and accounting systems that authorize and process transactions and provide the appropriate credits and debits to member accounts.

Getting both sides of the market on board requires that these associations provide members with the proper incentives to service both sides of the market and that there is a pricing structure that provides the proper incentives for individuals and merchants to use the cards. To see why, it is useful to go back to the case of American Express. As a company that both issues cards and acquires from merchants it can devise a pricing structure consisting of a price to cardholders (primarily an annual card fee) and a price to merchants (merchant discount). In contrast, the issuing and acquiring members of an association can only determine their own prices (card fees and interest charges in the case of issuing members and merchant discount and related fees in the case of acquiring members). They cannot determine the relative price of card services to individuals versus merchants; that must be done centrally.

The associations have done this through setting an "interchange fee." This is a fee that the bank that acquires a card transaction from a merchant charges the bank that issued the card to the individual who made that transaction. Issuing and acquiring banks had to have an agreement for consummating transactions. Both could take the position that the transaction could not take place without what they brought to the table (and both would be right). It was not practical for the thousands of members to negotiate individual agreements.[143]

The associations could have adopted a rule that said all transactions were exchanged at par—issuing banks had to fully reimburse acquiring banks. However, that in effect amounts to setting

---

[140] The benefit American Express cardholders receive from the float is estimated at approximately $32. This amount is calculated using average basic card member spending in 2001, a 28-day repayment grace period and the 2001 average 1-year constant maturity Treasury bill interest rate. *Id*.

[141] Including finance charge revenue in 2001, American Express earned 62 percent of its revenues from merchants. *Id*, at 35.

[142] For a short introduction, see David S. Evans, *Payment Card Business: Chickens and Eggs, and Other Conundrums,* Mastering Management Online (2001), in Fin. Times Online (visited Jul. 26, 2002) <http://www.ftmastering.com/mmo/mmo03_2.htm>.

[143] Section V reviews the origin of the interchange fee in the Visa association.

an interchange fee equal to zero and adopting a particular pricing structure. Estimates place the average Visa and MasterCard interchange fee in 2001 at 1.53 percent.[144] Acquiring banks usually pass that fee on to the merchant as either an explicit or implicit part of the merchant discount. The resulting pricing structure is one in which the average merchant discount for the card associations is substantially less than the merchant discount charged on similar cards issued by American Express and somewhat more than the merchant discount charged by Discover which is another integrated payment system.[145]

Payment cards are a complex match-maker market. The payment systems have a two-stage problem to solve. They have to persuade individuals and merchants to use the same platform—by getting individuals to carry the card brand and merchants to take the card brand. Many consumers do not carry American Express cards and many merchants do not accept those cards. They then need to persuade individuals to use the card to make purchases rather than competing payment devices and they need to persuade merchant to take the card for those purchases rather than steering consumers to other payment devices. Many consumers carry American Express cards but pay with a Visa cards or with cash for numerous transactions. Some merchants take American Express cards but discourage their use when the consumer has a payment alternative that is cheaper for the merchant. In effect, the systems want a lot of people to join the club and to go out on a lot of dates. The pricing structure, and investment strategy, seems to have arisen to solve these two problems: getting people on board and getting those on board together a lot.

Despite these differences with other two-sided markets there are many similarities. The payment systems have two groups of customers. The customers engage in transactions (the use of cards as a means of exchange) on a platform (the card systems). There are externalities—merchants value more cardholders; cardholders value more merchants. And there is a pricing structure that seems to skew prices towards one side—in this case the merchants.

III. ECONOMICS OF TWO-SIDED MARKETS

The preceding section cataloged a variety of business models and strategies employed by companies operating in two-sided markets. In this section, I explore the economic rationale behind the business models we observe in two-sided markets. Part A tries to isolate what makes two-sided markets different. It is true that they tend to have network effects and involve complementary products sold by multiproduct firms. But they have something more that makes the economics of them much different from the well-trodden fields of network economics and multiproduct pricing.

Part B presents a simple economic model of pricing in a two-sided market. It shows that the conditions for optimal product pricing are substantially different from those in the one-sided

[144] The Nilson Report No. 758 (Feb. 2002), at 4.
[145] See United States v. Visa, 163 F. Supp. 2d 322, 333 (S.D.N.Y. 2001).

markets usually analyzed by economists. Neither the usual price equals marginal cost condition found in perfectly competitive markets nor the usual marginal revenue equals marginal cost condition found in imperfectly competitive markets applies. Fundamentally, simple pricing and cost relationships cannot apply to either side because the transaction that takes place on the platform jointly benefits both parties and often entails joint costs.

Part C examines how the unique economics of two-sided markets affects the business models used by firms in these markets. It examines strategies for getting both sides on board, balancing the competing interests of both sides, and how these strategies are affected by the extent of competition among two-sided firms operating in the same market.

Part D concludes with a brief discussion of how the pricing strategies used in two-sided markets deviate from what a benevolent social planner, trying to maximize social welfare, would like to see. That will set up the discussion for the following section that explores the antitrust analysis of two-sided markets.

A. What Makes Two-Sided Markets Different

The difference between one-sided and two-side markets is subtler than it first appears. To see why, we need to distinguish two-sided markets from those with network effects and those involving complementary products. Network effects and complementarities are important aspects of two-sided markets but do not by themselves distinguish two-sided from one-sided markets.

1. Network Effects

A market has network effects (also known as network externalities or positive-feedback effects) when consumers value a product more the more other consumers use that product. In the case of direct network effects, I value (and therefore have a stronger demand for) the product because you have purchased it as well (therefore we can, for example, communicate with each other using this technology).[146] In the case of indirect network effects, I value (and therefore have a stronger demand for) the product because your purchase means that the demand for complementary products is higher and the supply of those complementary products will benefit me. Direct and indirect network effects result in purchasing decisions being interdependent over time.

All of the products we discussed in the previous section have indirect network effects. Bond sellers value an exchange system more if more bond buyers participate—they are more likely to consummate a transaction and obtain a more favorable deal. Merchants value a payment system

[146] One can take a communications network and think of there being two sides (senders and receivers) in which case one can also think of the network effects as indirect. What distinguishes two-sided from one-sided markets, and indirect from direct network effects, is whether one can identify distinct groups of users with different preferences. There is an argument that almost all network effects in reality are indirect effects and that almost all network markets are two-sided.

more if more customers have the card. Applications developers value an operating system more if more prospective applications buyers use it. Purchasers of video game consoles value them more if there are more users of these consoles because that will translate into more games. In all these cases, each side values having the other side on board and benefits the more customers there are on the other side.

2. Multiproduct Firms

All of the firms we discussed in the previous section sold what appeared to be multiple products. Residential agencies sell listing services as well as showing services, even though they are not typically priced separately. Magazines sell advertising space and content. Operating system vendors sell application developer tools and operating systems. Payment-card systems sell cardholder and merchant services. Many firms sell multiple products and there is an extensive economic literature on why they do so.[147] On the cost side, there may be economies of scope from having one firm produce multiple products. Automobile manufacturers can use the same production technology for making cars and trucks. American Express can use the same computer system for providing services to cardholders and merchants. On the demand side, there are advantages to pricing complementary products together.

That was first recognized by Cournot in 1838 and now goes by the unhelpful name of "double marginalization."[148] Suppose there are two complementary products, A and B. Two products are complements if the demand for one is higher the lower the price of the other. Unless there is perfect competition in products A and B, a firm that makes product A can increase its sales if the price of product B is lower (and vice versa for a firm that makes product B). By selling both products a firm can take this pricing interdependency into account; as it turns out, consumers generally benefit as the price of both complementary products will tend to be lower when they are priced jointly than when they are priced separately.[149] Gillette does this selling razors and razor blades. IBM does this by selling mainframes and peripheral devices. Microsoft does this by selling Microsoft Office and Microsoft Windows.

[147] Jeff Rohlfs, *A Theory of Interdependent Demand for a Communications Service*, 5 Bell J. Econ. 16 (1974); William J. Baumol *et al.*, Contestable Markets and The Theory Of Industry Structure (1982); John C. Panzar, *Technological Determinants of Firms and Industry Structure*, in Handbook of Industrial Organization Vol. I 3 (Richard Schmalensee & Robert D. Willig eds., 1989).

[148] Augustin A. Cournot, Researches Into The Mathematical Principles of the Theory of Wealth, (Nathaniel T. Bacon trans., Macmillan 1897).

[149] The classic example of this involves monopoly production of A and B. A merger of two monopolists in these complementary products would result in the merged firm charging lower prices for at least one of the products compared to the separate firm. *See generally* Thibaud Verge, *Multiproduct Monopolist and Full-Line Forcing: The Efficiency Argument Revisited*, 12 Econ. Bull. 1 (2001).

These standard explanations for why firms produce multiple products probably apply to many of the firms discussed in the previous section. But there is a further reason why firms sell multiple products in two-sided markets. Doing so helps them increase the indirect network effects discussed above. Most firms in one-sided markets do not have to offer product A in order to offer product B. GM could sell cars without trucks and Gillette could sell razor blades without selling razors. Many firms in two-sided markets, however, have to produce multiple products in order to sell any product at all. Diners Club could not have come into existence without providing different services to both cardholders and merchants. Selling operating systems and video game consoles requires firms in these industries to produce products for both users and developers. For others, producing multiple products is important but perhaps not essential. There are companies that specialize in sending advertising messages (direct mail, bill boards) and companies that specialize in just selling content to consumers (the leading German newspaper Frankfurter Allgemaine Zeitung has no ads, and Consumer Reports declines advertiser support).

B. Profit-Maximizing Pricing in Two-Sided Markets

A market is two-sided if at any point in time there are (a) two distinct groups of customers; (b) the value obtained by one kind of customers increases with the number of the other kind of customers; and (c) an intermediary is necessary for internalizing the externalities created by one group for the other group. Two-sided markets tend to result in businesses that supply both sides of the market, that adopt particular pricing and investment strategies to get both sides of the market on board, and that adopt particular pricing and product strategies to balance the interests of the two sides.[150]

A simplified model shows some of the key differences between one-sided and two-sided markets. I do the comparison under the assumption that a monopoly firm is providing the good. Firm S is in a single-sided market and sells q_S. Firm T is in a two-sided market and sells q_T^1 and q_T^2. I then explore the situation of competing firms.

1. Monopoly Two-Sided Firm

The pricing strategy of a monopoly provider in a one-sided market is well known. Assume that the firm S has a constant cost of production c_S. The monopoly charges price p_S and produces quantity q_S. Profits are: $\pi_S = (p_S - c_S) \times q_S$. To maximize profits the firm should increase q_S to the point where marginal revenue equals marginal cost. At this point the famous Lerner formula applies: percent markup of price over marginal cost should equal one divided by the elasticity of demand: $\dfrac{(p_S - c_S)}{p_S} = \dfrac{1}{\varepsilon}$, where the elasticity of demand is defined as the percentage change in

[150] I discuss what I mean by "balance the interests" below.

quantity resulting from a one percent change in price (times minus one so it is always positive).[151] If it were possible to identify separate groups of consumers with different demand the firm could engage in price discrimination in which case it would charge different prices to these different groups (where each price would be at the level at which the marginal revenue for that group equals marginal cost).

Rochet and Tirole examine the pricing and production strategy of a firm in a two-sided market that is motivated by payment cards but can be easily generalized to many two-sided match-maker markets.[152] They consider the case in which both sides of the market are buying a "transaction" and in which the seller incurs a marginal cost of c_T for consummating that transaction. The prices charged to buyers and sellers respectively are two prices, p_T^B and p_T^S. The buyer's demand $D_T^B\,(p_T^B)$ depends only on the price faced by the buyer and the seller's demand $D_T^S\,(p_T^S)$ depends only on the price faced by the seller. The demands can be thought of, roughly speaking, as the number of buyers and sellers using the system. The transactions that a seller engages in, and its benefits from those transactions, increase proportionally with the number of buyers on the system. The same holds for an individual buyer.[153]

Total demand equals the product of the two demands: $q_T = D_T^B\,(p_T^B)\,D_T^S\,(p_T^S)$. Thus, if there were 500 sellers and 100 buyers, there would be 50,000 transactions. This assumption mimics the situation in credit cards where the number of card transactions depends on the interactions between those who hold cards and those who take cards. It also captures the situation in which the number of video games that are played depends on the interactions between those who play games on a particular type of video game console and those who make games for that console. The two-sided monopoly's profits are: $\pi_T = (p_T^B + p_T^S - c_T) \times q_T$.

Although this model is special, the results described below are likely to hold generally: the firm has to choose a pricing structure in addition to a pricing level and profit-maximization does not result in the equilibration of marginal revenue and marginal cost in either market taken by itself. The assumption of multiplicative demand between the two sides understates the importance of the indirect network effects. It ignores the fact that the value each side obtains from the other side increases with the number of customers on the other side. The results below would likely be stronger if this feature were taken into account.[154]

---

[151] See Carlton & Perloff, *supra* note 6, at 91–92. (The elasticity is sometimes defined without multiplying by minus one.)

[152] *See* Rochet & Tirole, *supra* note 4.

[153] Rochet and Tirole assume, for simplicity, that there are no fixed costs of joining the system. If there were fixed costs, a buyer would have to weigh the overall gain from joining, which would depend on the number of sellers, against the fixed costs, so that the number of buyers joining would depend on the number of sellers. As a result, the decision of an individual seller to join the system does not depend on the number of buyers.

[154] The particular model is used because it leads to tractable results that can be used to analyze the welfare consequences of two-sided pricing.

A key difference between the one-sided and two-sided firms is that, as Rochet and Tirole show, the two-sided monopolist must choose a pricing level (what total price to charge to buyers and sellers) and a pricing structure (how to divide the total price between buyers and sellers). The pricing level condition is: T $\frac{(p_T^B + p_T^S) - c_T}{(p_T^B + p_T^S)} = \frac{1}{\varepsilon^B + \varepsilon^S}$. This condition is a variant of the Lerner condition above. Here however, analog of the price term in the one-sided Lerner condition is the sum of the prices to both sides and the analog of the demand elasticity term is the sum of the demand elasticities on both sides.

The optimal pricing structure depends on the following condition where the D' refers to the change in demand with respect to a change in the price on that side of the market: $(D_T^B)'D_T^S = D_T^{B}{}'(D_T^S)$. Consider the left-hand side of this equality. A slight increase in price on the seller's side will decrease the seller side demand slightly, by $(D_T^B)'$. The impact on total demand (and profits) is the product of the seller side effect times the buyer side demand. The right hand side is the analogous effect of increasing the buyer side price slightly. In equilibrium, the effect on profits must be the same from increasing the seller side price versus the buyer side price. That implies $\frac{p_T^B}{\varepsilon^B} = \frac{p_T^S}{\varepsilon^S}$, where the ε's are the elasticities of demand for each side of the market.

This is very different from the Lerner condition, shown above, in one-sided markets. Moreover, none of the conditions for determining the price level or the price structure in two-sided markets corresponds to marginal revenue equaling marginal cost on either side of the market. In fact, such conditions have no meaning in two-sided markets because there is no way to allocate the increases in revenues from changes in prices to one side or the other. Changes in prices result in more "transactions" from which each side jointly benefits. Nor is there any way to allocate the costs. Often costs are jointly incurred for both parties to a transaction and we have the usual issue that any allocation of cost is arbitrary.[155]

To try to understand the pricing structure condition intuitively, consider the following. Suppose, for example, the seller side demand (D_T^S) is five times larger in equilibrium than the buyer side (D_T^B)—say, 500 to 100, with total output of 50,000. Since total demand is the product of the demands on the two sides, a change in the buyer side demand has five times the effect on total demand as the same sized change in seller side demand. If the buyer side decreases by one unit, to 99, then total output falls by 500 units. But if the seller side decreases by one unit, total output falls by only 100 units. At equilibrium, the change in profits coming from a one-cent increase in the seller price must be the same as from the same increase in the buyer price. The monopolist therefore wants to set prices so that the change in buyer side demand from a marginal price change is only 1/5 of the change in seller side demand from a marginal price change because the effect of a buyer side change is multiplied against the much larger demand on the seller side.

[155] *See supra* note 29.

2. Competing Two-Sided Firms

The results are broadly similar when there are competing firms selling to both sides of the market. In practice, consumers in two-sided markets tend to engage in "multihoming"[156]—consumers on one or both sides of the market rely on more than one seller of two-sided services. For example, game developers write for several different consoles, merchants accept several different types of credit cards, and home buyers sometimes use several different real-estate agents. Rochet and Tirole consider multihoming under the further assumption—often true in practice—that one side of the market can dictate which two-sided firm must be used in any particular transaction. Payment cardholders usually decide which card to use and computer users usually decide which operating system to use. In this case, competing two-sided firms still must choose a pricing level and a pricing structure. However, the relevant demand elasticities are increased by a factor that reflects the extent to which consumers multihome and therefore have substitutes readily available.

C. Business Models in Two-Sided Markets

Although the economics presented above is simple it helps us understand the rationale for the business models that have been adopted in two-sided markets. Here I consider several issues that occur repeatedly in two-sided markets.

1. Getting both sides on board. There are many references in the literature on the firms discussed earlier about solving the chicken-and-egg problem.[157] For example, there would be no demand by households for payment cards if they could not use them anywhere and no demand by retailers for payment cards if no one had them. Which comes first—the cardholder or the retailer?[158] Investment and pricing strategies are key to getting both sides on board.
2. Balancing interests. Even with both sides on board, businesses have to carefully balance their two demands. They always have to consider how changing prices on one side of the market will impact the other side of the market. Businesses sometimes differ over this. For example, in the battle between Microsoft and Netscape over Internet browsers, Microsoft

[156] "Multihomed" was originally an Internet term. According to Webopedia, an online technical dictionary, it is "an adjective used to describe a host that is connected to two or more networks or has two or more network addresses. For example, a network server may be connected to a serial line and a LAN or to multiple LANs." For definition, see Webopedia Online Dictionary (visited Aug. 13, 2002) <http://www.pcwebopaedia.com/TERM/m/multihomed.html>. Rochet and Tirole adapt the term to describe two-sided networks where a fraction of end-users on one or more sides connect to multiple platforms. See Rochet & Tirole, *supra* note 4, at 5.

[157] *See*, e.g., Gawer & Cusumano, *supra* note 28, at 150-151, 60-61.

[158] *See* Evans & Schmalensee, *supra* note 10, at 137-168.

gave away developer kits to Internet portals while Netscape charged for them.[159] Political tensions can also manifest themselves—looking out for their narrow interests customers on each side of the market would like the other side to pay more. This is a familiar problem in the payment-card industry—in Europe a retailers association asked the European Commission to force the card associations to eliminate the interchange fee.[160]

3. Multihoming. Firms sometimes compete to become the dominant two-sided provider. A monopoly provider can emerge if consumers are sufficiently homogeneous (so that one firm can provide them all), the indirect network effects are strong enough, and there are sufficiently large scale economies in production. But even in these circumstances, at the start of an industry it is common to see several firms in a race to become the dominant firm—they are likely to engage in multihoming strategies. If some cases it is possible that several two-sided firms could co-exist and compete with each other—this has clearly been the case in payment cards where there have been multiple providers for almost half a century. In this case, multihoming has implications for how they compete with each other.

1. Getting Both Sides on Board

An important characteristic of two-sided markets is that the demand on each side vanishes if there is no demand on the other—regardless of what the price is. Men will not go to dating clubs that women do not attend because they cannot get a date. Merchants will not take a payment card if no customer carries it because no transaction will materialize. Computer users will not use an operating system that does not have applications they need to run. Sellers of corporate bonds will not use a trading mechanism that does not have any buyers. In all these cases, the businesses that participate in these industries have to figure out ways to get both sides on board.

One way to do this is to obtain a critical mass of users on one side of the market by giving them the service for free or even paying them to take it. Especially at the entry phase of firms in two-sided markets it is not surprising to see precisely this strategy. Diners Club gave its charge card away to cardholders at first—there was no annual fee and users got the benefit of the float. Netscape gave away its browser to most users to get a critical mass on the computer user side of the market; after Microsoft started giving away its browser to all users Netscape followed

[159] Cameron D. Myhrvold, *Microsoft's Efforts to Create Alliances with ISPs,* Microsoft PressPass, Jan. 6, 1999, (visited Aug. 21, 2002) <http://www.microsoft.com/presspass/trial/mswitness/myhrvold/myhrvold_pt2.asp>.

[160] See *Retail Disappointment as Visa Exempted from EU Competition Rules*, British Retail Consortium Gruerrera, *Brussels Set to Let Visa off the Hook*, Fin. Times, May 13, 2002; James MacKintosh, *Watchdog Rules Credit Card Fees to Retailers Excessive*, Fin. Times, Sept. 26, 2001.

[161] Wylie Wong, *Netscape Applauds Microsoft Suit*, Techweb, May 20, 1998 (visited Aug. 21, 2002) <http://www.techweb.com/wire/story/msftdoj/TWB19980519S0007>.

suit.[161] Microsoft is reportedly subsidizing the sales of its X-box hardware to consumers to get them on board.[162]

Another way to solve the chicken-and-egg problem is to invest in one side of the market to lower the costs to consumers on that side of participating in the market. Microsoft provides a good example of this. As we saw earlier it invests in applications developers by developing tools that help them write applications and providing other assistance that makes it easier for developers to write applications using Microsoft operating systems. To take another example, bond dealers take positions in their personal accounts for certain bonds they trade. They do this when the bond is thinly traded and the long time delays between buys and sells would hinder the market's pricing and/or liquidity. By investing in this manner, two-sided intermediaries are able to cultivate (or even initially supply) one side, or both sides, of their market in order to boost the overall success of the platform.

Providing low prices or transfers to one side of the market helps the platform solve the chicken-and-egg problem by encouraging the benefited group's participation––which in turn, due to network effects, encourages the non-benefited group's participation. Bernard Caillaud and Bruno Jullien refer to this strategy as "divide-and-conquer."[163] Another effect of providing benefits to one side is that this assistance can discourage use of competing two-sided firms. For example, when Palm provides free tools and support to PDA applications software developers, it encourages those developers to write programs that work on the Palm OS platform, but it also induces those developers to spend less time writing programs for other operating systems.[164]

2. *Pricing Strategies and Balancing Interests*

Firms in mature two-sided markets—i.e. those that have already gone through the entry phase in which the focus is on solving the chicken-and-egg problem—still have to devise and maintain an optimal pricing structure. In most observed two-sided markets, companies seem to settle on pricing structures that are heavily skewed towards one side of the market. Table 1 in Appendix A summarizes the pricing structure for the markets we have identified as two-sided. For example, in 2001, excluding finance charge revenue American Express earned 82 percent of its revenues from merchants.[165] Microsoft earns the substantial majority of its revenue from Windows from licensing

[162] David Becker, *Xbox Drags on Microsoft Profit*, CNET.COM, Jan. 18, 2002 (visited Aug. 21, 2002) <http://news.com.com/2100-1040-818798.html>.

[163] *See* Caillaud & Jullien, *supra* note 4, at 16. *See* also Jullien, supra note 4, at 1.

[164] *See* Rochet & Tirole, *supra* note 4, at 6.

[165] If finance charge revenues are included, American Express earned 62 percent of its revenues from merchants in 2001. *See* American Express Company Annual Report 2001, *supra* note 138, at 35.

Windows to computer manufacturers or end-users.[166] Real estate brokers usually earn most or all of their revenues from the sellers.

Discerning the optimal pricing structure is one of the challenges of competing in a two-sided market. Sometimes all the platforms converge on the same pricing strategy. Microsoft, Apple, IBM, Palm and other operating system companies could have charged higher fees to applications developers and lower fees to end-users. They all discovered that it made sense to charge developers relatively modest fees for developer kits and, especially in the case of Microsoft, to give a lot away for free. Nevertheless, Microsoft is known for putting far more effort into the developer side of the business than the other operating system companies.[167]

The debit card is an example in which different platforms made different pricing choices. In the late 1980s, the ATM networks had a base of cardholders who used their cards to withdraw cash or obtain other services at ATMs. They had no merchants that took these cards. To add debit services to existing ATM cards, the ATM networks charged a small interchange fee (8 cents per transaction on a typical $30 transaction) to encourage merchants to install pin-pads that could read the ATM cards that cardholders already had and accept the pins they used to access the ATM machines.[168] Many merchants invested in the pin-pads—the number of pin-pads increased from 53,000 in 1990 to about 3.6 million in 2001.[169] The credit-card systems had a base of merchants who took their cards but it did not have cards that, like the ATM cards, accessed consumers' checking accounts. The credit-card systems imposed a much higher interchange fee than the ATM networks, about 38 cents versus 8 cents on a typical $30 transaction.[170] They did this to persuade banks to issue debit cards and cardholders to take these cards.[171] The number of Visa debit cards in circulation increased from 7.6 million in 1990 to about 117 million in 2001.[172]

---

[166] From 1988 through 2000, Microsoft earned at least 67 percent of its revenues from licensing packaged software (such as Windows and Office) to end-users (either directly at retail or through manufacturer preinstallation on PCs). See *1994 Worldwide Software Review and Forecast*, IDC REPORT #9358 (Nov. 1994); *1995 Worldwide Software Review and Forecast*, IDC REPORT #10460 (Nov. 1995); *1996 Worldwide Software Review and Forecast*, IDC REPORT #12408 (Nov. 1996); *1997 Worldwide Software Review and Forecast*, IDC REPORT #14327 (Oct. 1997); *1999 Worldwide Software Review and Forecast*, IDC REPORT #20161 (Oct. 1999); *Worldwide Software Market Forecast Summary, 2001-2005*, IDC REPORT #25569 (Sept. 2001). Note that the 67 percent figure underestimates the amount of revenue Microsoft earns from end-users because the other third of revenue coming from "Applications Development and Deployment" includes some end-user revenues as well. For example, database products used by business IT departments are included in the Applications Development category.

[167] *See* GAWER & CUSUMANO, *supra* note 28, at 150-151.

[168] See EVANS & SCHMALENSEE, *supra* note 10, at 300.

[169] *Id.* at 308-309; and THE NILSON REPORT No. 759 (March 2002), at 6.

[170] The ATM systems typically charged a flat interchange fee per transaction, while the interchange fee set by Visa and MasterCard varied with the size of the transaction. The reported interchange fee comparison is from 1998, around the time of substantial growth in debit for the ATM and credit-card systems. *Id.* at 300.

[171] Visa attracted consumers through an effective advertising campaign and attracted issuers through heavy investment in a debit processing facility, among other strategies. *Id.* at chapter 12.

[172] *See* THE NILSON REPORT No. 760 (March 2002), at 7; THE NILSON REPORT No. 500 (May 1991), at 6.

Two other factors influence the pricing structure. There may be certain customers on one side of the market—Rochet and Tirole refer to them as "marquee buyers"—that are extremely valuable to customers on the other side of the market. The existence of marquee buyers tends to reduce the price to all buyers and increases it to sellers. A similar phenomenon occurs when certain customers are extremely loyal to the two-sided firm—perhaps because of long-term contracts or sunk-cost investments. For example, American Express has been able to charge a relatively high merchant discount as compared to other card brands, especially for their corporate card, because merchants viewed the American Express business clientele as extremely attractive. Corporate expense clients were "marquee" customers that allowed American Express to raise its prices to the other side of the market, merchants. In the online debit-card market, however, card issuers faced "captive" customers—ATM cards could be used as online debit cards, so consumers did not need to be courted to accept the new payment form. Therefore, it has been the merchants—who must purchase and install expensive machinery in order to process online debit transactions—who have been courted, as we saw above.

3. Multihoming

Most two-sided markets we observe in the real world appear to have several competing two-sided firms and at least one side appears to multihome. Table 2 in Appendix A presents a summary. Consider, for example, personal computers. One could consider the two sides as consisting of personal computer end-users and as developers of applications. The end-users do not multihome. They almost always use a single operating system and by far the preponderance of them use a Microsoft operating system.[173] The developers do multihome. According to Josh Lerner, in 2000, 68 percent of software firms developed software for Windows operating systems, 19 percent for Apple computers operating systems, 48 percent for Unix operating systems including Linux, and 36 percent and 34 percent for proprietary non-Unix operating systems that run on minicomputers and proprietary operating systems that run on mainframes respectively.[174] In fact, in recent years the percentage of software firms developing for non-Microsoft operating systems has increased. The fastest-growing category has been software firms developing for Unix operating systems including Linux. The percentage of developers in this category increased from 29 percent in 1998 to 48 percent in 2000.[175]

[173] *Worldwide Operating Environments Market Forecast and Analysis Summary, 2001-2005*, IDC Report #25119 (Aug. 2001), at 12.

[174] *See* Josh Lerner, *Did Microsoft Deter Software Innovation?* Working Paper (Jan. 2002) (downloaded Aug. 15, 2002) <http://gsbwww.uchicago.edu/research/workshops/elo/lerner2.pdf>; and Corporate Technology Directory, Editions 1990-2000. The percentages total to 205, indicating substantial multihoming on the part of developers.

[175] *Id.*

Multihoming affects both the price level and the pricing structure. Not surprisingly the price level tends to be lower with multihoming—the availability of substitutes tends to put pressure on the two-sided firms to lower their prices. The seller has more options dealing with a multihomed buyer on the other side and can steer toward its preferred platform. As buyer multihoming becomes more prevalent, prices to sellers will tend to decrease since they have more substitution options.

Even when multihoming is not prevalent on one side of a two-sided market the possibility of multihoming may have significant consequences for pricing. The possibility of multihoming may encourage firms to lower their prices on the side of the market in which multihoming could occur. By lowering their prices they discourage customers on that side from affiliating with other two-sided firms. This is not entirely a free lunch for consumers. The firm can then charge more to customers on the other side, for whom fewer substitutes are available.[176]

D. Two-Sided Markets and Social Welfare

A relatively small number of firms tend to compete in two-sided markets. That is because these markets have network effects and usually incur substantial fixed costs for getting one or both sides on board. Larger firms have advantages over smaller firms, at least up to a point, because their larger size delivers more value—a bigger network—to consumers. In the case of two-sided markets, larger firms are able to deliver a bigger network of customers on one side of the market to customers on the other side of the market. As we have seen, other economic factors, in particular the existence of heterogeneous consumers on one side of the market or the other, tend to limit the importance of network effects so that it is possible for multiple firms to compete in two-sided markets.

Firms in concentrated two-sided markets, like firms in all concentrated markets, may have opportunities to earn supra-competitive profits—i.e. profits that exceed those necessary to attract capital to the industry after accounting for risk. Several factors affect the extent to which this can happen over time.

1. The extent to which firms are competing to become established in a two-sided market. Firms tend to compete to establish customer bases on both sides of the market. This results in investments to court customers, to provide them with subsidies in the form of equipment, and to offer them low or negative prices.[177] If the competition is sufficiently intense

[176] In Jullien's model, when multiple platforms compete and price discrimination between the two customer types is possible, then prices are lower overall: "This forces the established firm to set on average prices at a much lower level than it would do with uniform prices. It turns out that it is impossible for a network to capture in equilibrium the surplus generated by the inter-group network externalities." Jullien assumes the incumbent initially offers uniform prices because in his model the two customer types have identical valuations for the network goods and both receive the same extra value if they both join the same network. *See* Jullien, *supra* note 4, at 4.

[177] Losing money initially to buy penetration can also be an important phenomenon in one-sided networks.

the losses incurred during the "getting both sides on board" stage may significantly offset the profits earned during the mature phase of the industry. For example, firms that have entered the payment-card industry have all incurred sizeable losses during their startup phases.[178]

2. The extent to which there are first-mover advantages in getting either side of the market on board. If any firm can make investments to get a side of the market on board then competition to make these investments should reduce the opportunities to earn significant supra-competitive returns. Supra-competitive profits could arise if one firm has an advantage that other firms cannot replicate. For example, the network economics literature argues that the first mover in an industry always has an advantage over the subsequent movers.[179] Since customers value other customers whoever gets customers first naturally wins.[180] It could also be that a firm has some asset that gives it a hard start. For example, a firm may have already developed a product that gives it a large customer base on one side of a market. When the demand for a two-sided version of this (or a related) product appears it will have a significant head start over rivals. Online debit is an example. Banks issued ATM cards to customers for use at its ATM machines. The ATM networks they belonged to eventually realized that these same ATM cards could be used to pay at merchants. The fact that they had the cardholder side on board gave them a significant advantage.

3. Even markets that appear to be dominated by a single player may in reality be contestable. Jullien's "model suggests that it may be easier than expected for a superior technology to enter, provided that the quality improvement is large enough."[181] Because many of the two-sided markets are fast moving, current leaders often face considerable competition in the form of potential entrants—other platforms striving to displace today's leader. Caillaud and Jullien argue that the Internet represents one such environment: "Too many ways of stealing the competitors' business appear. Unsurprisingly, the strategic situation is very unstable and the only equilibrium situation that is tenable is for a firm to exert dominance on the intermediation market, i.e. to be the sole supplier of intermediation services, without enjoying any market power as potential entrants create a strong disciplinary device for the dominant firm. In some sense, this market is extremely contestable."[182]

---

[178] *See* Evans & Schmalensee, *supra* note 10, at ch. 4.
[179] *See*, e.g., Carl Shapiro & Hal R. Varian, Information Rules 29-32 (1999).
[180] There is little empirical support, however, for this view. *See* Evans & Schmalensee, *supra* note 7. See *also* Stan J. Liebowitz & Stephen E. Margolis, Winners, Losers & Microsoft: Competition and Antitrust In High Technology (1999).
[181] *See* Jullien, *supra* note 27, at 34.
[182] *See* Caillaud & Jullien, *supra* note 4, at 40. The authors are speaking of Internet intermediaries, but the point holds for other fast-moving dynamic markets.

4. Two-sided markets in which non-profit associations determine the pricing structure are not likely to permit the participants to earn supracompetive profits. Payment-card associations have put an effectively non-profit institution in charge of managing a physical network for members and for determining pricing policies. Pricing levels are determined by competition among members of the association. For example, in the United States payment-card industry, thousands of banks compete for cardholders; although a small number of firms compete for merchant services this business is considered to be highly competitive also. Interest rates, card fees, merchant charges, and so forth are determined through intense competition. Pricing structures are affected by the non-profit institution. In the case of payment cards, the interchange fee affects the relative price to cardholders and merchants.

The consequences of having relatively few competitors in two-sided markets, and the existence of network effects, raise familiar issues concerning the efficacy of competitive markets and the possible roles for government intervention. The pricing and investment strategies that firms in two-sided markets use to "get both sides on board" and "balance the interests of both sides" raise novel ones. These pricing and other business strategies are needed to solve a fundamental economic problem arising from the interdependency of demand on both sides of the market. In some cases, the product could not even exist without efforts to subsidize one side of the market or the other.

Rochet and Tirole, in an admittedly simplified setting, have compared the pricing structure adopted by firms in two-sided markets to the pricing structure that would maximize social welfare. Interestingly, they find that a monopoly firm, a firm with competition, and a benevolent social planner would adopt similar pricing structures. The precise relative prices would differ somewhat.[183] However, Rochet and Tirole find that the pricing structure adopted by the market (monopoly or two-firm oligopoly) is not biased towards one side of the market or the other side of the market compared to the pricing structure that would be adopted by a benevolent social planner.[184] (Schmalensee finds similar results for interchange fees.) [185] Therefore, there is no reason to believe that the preferences that one side or the other of the market have gotten, as summarized in Table 1, are systematically different from what a social planner would seek to achieve.

IV. ANTITRUST ISSUES RAISED BY TWO-SIDED MARKETS

The economics of two-sided markets differ from the economics of one-sided markets in important respects. First, the individual prices charged on either side of the market do not track costs or demand on that side of the market. Indeed, the fact that benefits and costs arise jointly in the two

[183] In the very special case of linear demand the pricing structures would be identical. See Rochet & Tirole, *supra* note 4, at 25.
[184] *Id.* at 24.
[185] *See* Schmalensee (2002), *supra* note 4, at 118-120.

sides of the market means that there is no meaningful economic relationship between benefits and costs on either side of the market considered by itself. It takes two to tango. Second, one cannot talk about the individual prices in isolation. Any change in demand or cost on either side of the market will necessarily affect both prices along with the sum of those prices. Third, products in two-sided markets cannot come into existence and cannot remain in existence unless firms in those markets get "both sides on board." This gives rise to pricing and investment strategies that differ from those taken in one-sided markets and seem odd unless considered in the context of competition in a two-sided market. Fourth, any analysis of social welfare must account for the pricing level, the pricing structure, and the feasible alternatives for getting both sides on board. It must also account for the extent to which not-for-profit institutions manage those aspects of the network that could give rise to supra-competitive profits.

These differences matter for antitrust analysis. To see the principles consider the following hypothetical merger. There are two chains of dating clubs in Tokyo—AAA Mates and Best Match (clubs A and B, respectively). They cater to somewhat different clienteles. Club A charges men $20 for admission and women $0; Club B charges men $30 for admission and gives women a $5 credit (in the form of free drinks). Club B has a 40 percent market share while Club A has a 10 percent market share. Club B has been more successful because it attracts more women and as a result of that it attracts more men. In fact, it is so successful that—like an "in" discotheque—it typically has a line and can select the men and women to admit. It tries to weed out "undesirable" men and women.

In analyzing whether this merger would be bad for consumers, a competition regulator would have to take several factors into account. Let us start by considering market definition. A preliminary issue is whether dating clubs is the relevant antitrust market in which this merger should be evaluated. Under the U.S. merger guidelines, it would be a market if a merger of all the dating clubs in Tokyo (assume Tokyo is the relevant geographic market) would enable the combined firm to raise price by a small but significant non-transitory amount—let us say 10 percent. For analyzing that question one should look at the "total price" charged by the club—although the average price across all patrons of the club would provide the same information. There is no reason to focus on the individual prices for men and women (except perhaps as they say something about product differentiation among the clubs). Different competitors may choose different pricing structures depending on the mix of men and women they want to attract. They could easily adjust these pricing structures in response to a change in the market.

If dating clubs is a relevant antitrust market, the next issue is whether a merger of clubs A and B would result in prices increasing. One cannot answer that question by looking just at the demand for patrons overall—e.g. by estimating the demand for admission against the average price. The mix of men and women is critically important. One would have to estimate the demand for men and the demand for women simultaneously. Then, using the theory of pricing in two-sided markets considered earlier together with information on cost, one could predict whether the merger would lead the combined firms to increase their total price.

Let us suppose that the competition authority has completed this exercise and found that the merged dating clubs would charge $32 for men and give women a credit of $6. Assuming equal numbers of men and women, the average price charged at Club A would rise from $10 to $13 and the average price charged at Club B would rise from $12.50 to $13. Should the competition authority ban the merger based on this price increase? From the standpoint of consumers the answer is unclear. At this point in a merger investigation the companies could come back and argue that there are efficiencies from the merger and that the reductions in cost will pull prices back down. There are no such savings here. But there is something else. Club B may have efficiencies that enable it to provide a higher value to consumers on the two sides of the market. On average the customers pay more but the men may have a better selection of women to choose from and the women may have a better selection of men to choose from.[186] In effect, there are demand-side efficiencies that result from increasing the indirect network effects realized by both sides of the market.

The remainder of this section examines how the economics of two-sided markets affects several key areas of antitrust analysis. Part A considers market definition. It focuses on how the merger guidelines approach should be modified for mergers of firms in two-sided markets. Part B examines the analysis of market power—the ability of firms to hold prices significantly above marginal cost. It considers the implications of the breakdown of the usual price equals marginal cost condition in two-sided markets. Part C considers the analysis of barriers to entry in two-sided markets—this is an important topic that comes up in both market definition and market power. Part D looks at efficiency considerations for two-sided markets; as part of this exercise it examines how the two-sided nature of markets should affect the classification of practices as per se versus rule of reason. Part E examines the implication of getting and keeping both sides on board in two-sided markets for the analysis of predation. Part F looks at a variety of other issues.

A. Market Definition

The general purpose of market definition is to provide a context for examining the issues that arise in an antitrust matter. For merger cases market definition helps identify the firms that could constrain possible price increases by the merging parties. For cases involving alleged anticompetitive behavior, market definition helps determine whether the defendant has enough market power for its behavior to have ultimately harmful effects on competition and consumers. Often market definition results in determining whether a firm is "in" the market or "out of" the market and then calculating the shares of the firms "in the market." A high share indicates market power or monopoly power.

In the case of mergers, the U.S. Department of Justice and Federal Trade Commission have developed techniques for determining whether a firm is in the market. They start with the firm(s) under consideration and add competitors to the market. The market boundary results (in a geo-

[186] Merger analysis in one-sided markets faces similar problems. Sometimes mergers permit the parties to create new products. The value of these new products should be considered as part of the efficiency analysis.

graphic or product dimension) when the collection of firms could, acting as a monopolist, raise price by a small but significant non-transitory amount (often taken to be 5-10 percent). If the collection of firms could do so, then presumably the firms "outside of the market" do not constrain the firms "inside the market" much.

For two-sided markets, this analysis must pay attention to both sides. It should consider firms that currently supply both sets of customers. Here the main controversy is likely to be over whether the two-sided firms offer sufficiently differentiated products that they should not be considered in the same market. Dating Club B might be the uptown/yuppie club and Dating Club A might be the downtown/blue collar club. To take another example, one of the issues in litigation concerning credit cards is whether other two-sided products such debit cards, checks, and cash compete with credit cards. The market analysis should also consider the possibility that other firms could enter. Dating Club A may be able to enter the uptown/yuppie market rather easily. During the early years of Palm PDAs it would have been reasonable to argue that Microsoft could easily enter this operating system category—in fact it eventually did so by producing and licensing a small operating system to a PDA manufactures.[187] Of course, in considering the possibility of entry it is necessary to examine whether and to what extent there are barriers to entry, a topic we consider below.

The right question to ask, for the merger guidelines price test, is whether the firms under consideration, if they merged today and priced as a monopolist, could raise the total price by 10 percent or more. There is no particular reason to focus on price effects on one side or the other—both groups of consumers matter. That is especially true in matching markets in which both sides must buy for there even to be a product. Moreover, competition authorities should not consider a price increase of 10 percent to one side of a two-sided market (in which the other side has no increase) as seriously as a price increase of 10 percent price to a one-sided market. Likewise, they should not ignore any countervailing price decreases that occur on the other side of the market.

Generally, it is not possible to examine price effects on one side of a market without considering the effect on the other side and the feedback effects between them.[188] For example, holding the female side of the market constant (price and quantity), perhaps the merger of dating clubs would result in an increase in the price to the male side of the market by 10 percent. But in reality it is not possible to hold the female side of the market constant. A 10 percent price increase to men would reduce the number of male patrons and therefore reduce the value of the service to women. Once these interdependencies are taken into account it could be that the merged firms could raise combined prices by only a trivial amount.

---

[187] *See* Microsoft Corp. website (visited Aug. 20, 2002) <http://www.microsoft.com/windows/embedded/>.

[188] Ignoring interdependencies between the two sides is a common mistake in the antitrust analysis of two-sided markets. For example, this mistake was made in the analysis of class certification issues in *Visa Check/MasterMoney. See In re* Visa Check/MasterMoney Antitrust Litig.; Wal-Mart, *et al*. v. Visa U.S.A. Inc. and MasterCard International, Inc., 280 F.3d 124 (2d Cir. 2001). The plaintiffs' expert claimed that it was possible to change a significant price on one side of the market (debit-card interchange fees) without any other changes in the two-sided market. For more details, see David S. Evans, *Class Certification, the Merits, and Expert Evidence,* GEO. MASON L. REV. (forthcoming).

The Justice Department's approach in *U.S. v. Visa U.S.A. et al.*[189] illustrates the mistake. MasterCard and Visa service cardholders and merchants. The DOJ's economic expert asked the question of whether a hypothetical merger of all credit and charge card issuers could profitably raise prices to cardholders, looking only at profits on the issuer/cardholder side.[190] He thus failed to consider two important factors. First, any decrease in cardholder volume would necessarily and directly lead to a decrease in merchant volume. And if merchant volume decreases, then any profits on the merchant side also decrease. DOJ's economist did not consider effects on profits on the merchant side. And second, a decrease in the cardholder base makes the system less attractive for merchants, thus potentially leading to a decrease in merchant demand for the system. (Which could then lead to a decrease in cardholder demand, and so on.) These changes, not accounted for by the government's analysis, would affect profits on both the issuing and acquiring sides. By focusing only on the cardholder side, the analysis put forward by the government's economist neglected at least half of the story.

B. Market Power

Taking both sides of the market into account is also important for analyzing market power. A traditional question is whether the firm under consideration prices above marginal cost by a significant amount. We saw earlier that there is no necessary relationship between price and marginal cost on either side of the market. In fact, the price on one side of the market could be well above marginal cost while the price on the other side of the market could be below marginal cost. To analyze market power one therefore has to examine whether the total price is significantly above marginal costs. For example, in the residential real estate market, agents must expend time and effort each time an open house is held for a seller and each time a buyer is taken to visit another property. Assuming only one agent is involved, the agents' 6 percent commission may well exceed some conception of the costs incurred directly for the seller (although, as we have discussed, all costs are really incurred for both sides), but the potentially numerous appointments with home buyers must be considered as well. And, as is true in one-sided markets, the risk of making no sale at all must also be considered.

Of course, in markets in which there are significant fixed costs, looking at price-cost margins is never a meaningful measure of market power. In those cases it makes more economic sense, in theory, to look at the risk-adjusted rate of return on investment. (Unfortunately, in practice it is extremely difficult to determine whether a firm or an industry—one-sided or two-sided—earns a supra-competitive risk-adjusted rate of return.) [191] For two-sided markets that analysis should consider the total returns and the total investment in both sides. In the video game market, while the cost of produc-

[189] *See* United States v. Visa, 163 F. Supp. 2d 322 (S.D.N.Y. 2001).
[190] *Id.* at 336.
[191] *See* Franklin M. Fisher & John J. McGowan, *On the Misuse of Accounting Rates of Return to Infer Monopoly Profits,* 73 Am. Econ. REV. 82 (1983) for the seminal treatment of this issue which turns on the fact that it is difficult to measure ex post the expected return *ex ante*.

ing one more CD copy of an existing game is trivial, considerable investment is required to initially develop that game. Platforms, such as Sony PlayStation, need to price all of the components—the console, the internally developed games, and the licensing fees for externally developed games—such that it recovers its R&D expenses at the same time as it maximizes the popularity of the platform.

C. Barriers to Entry

The existence of barriers to entry comes up in both market definition and market power analyses and deserves a separate treatment. In market definition, barriers to entry come up in determining whether it is possible for other competitors to come into the market. In market power, barriers to entry come up in determining whether the defendant can exclude competitors and thereby maintain higher prices. This is of particular concern in monopoly maintenance cases where a preliminary issue is whether the defendant has monopoly power. According to *du Pont*, a firm has monopoly power if it has the power to "control prices or exclude competition."[192]

There is some disagreement among economists and the courts over the proper definition of a barrier to entry. Consider the following example. A firm spent $100 million to build a plant to make widgets. There is nothing special about this plant—anyone could build a similar plant for $100 million. Suppose there are two widget firms that have built two plants each costing $100 million. Is there a barrier to entry into the market? Some economists and courts have used barriers to entry in the colloquial sense—it is hard to get into a market because it costs money and effort.[193] They would agree that there is a barrier to entry into the widget market.

[192] United States v. E.I. du Pont, 351 U.S. 377, 391 (1956).

[193] Phillip E. Areeda *et al.*, Antitrust Law, Volume IIa: An Analysis of Antitrust Principles and Their Application 55-56 (1994). "A barrier to entry is any factor that permits firms already in the market to earn returns above the competitive level while deterring outsiders from entering" (footnote omitted). This definition follows J.S. Bain, Barriers To New Competition: Their Character and Consequences in Manufacturing Industries (1962). For a list of court cases relying on this definition of entry barriers, see Phillip E. Areeda *et al.,* Volume IIA: An Analysis of Antitrust Principles and Their Application 123-124 (Supp. 1999). In practice, the condition that the factor must result in the ability of the firm "to earn returns above the competitive level" is dropped in one of two ways. First, the condition is just ignored. The correct economic concept is whether the expected rate of return to the firm based on the information available to the market at the time it made the investment exceeds the competitive level after adjusting for risk. See Fisher & McGowan, supra note 191, at 90-91. It is almost impossible to measure this *ex ante* return with ex post information. Second, the competitive rates of return are calculated from *ex post* data with no adjustment for the risk perceived *ex ante*. For example, it is common for analysts to calculate the returns based on survivors in an industry without adjusting for the losses incurred by failures. The result is that almost all industries involving substantial investments in research and development or other fixed costs are identified, incorrectly, as earning a supra-competitive return. See Fisher & McGowan, *supra* note 191, at 91; and Franklin M. Fisher *et al,* Folded, Spindled and Mutilated: Economic Analysis And U.S. V. Ibm ch. 7 (1983). Not surprisingly, Areeda, Hovenkamp and Solow identify almost any advantage of incumbency as a barrier to entry. They include "economies of scale, high initial investment, capital market imperfections, risk, low prices, scarce inputs or customers, product reputation and promotion, and government constraints" as barriers to entry. Areeda *et al.* (1994), supra, at 63.

Many economists have adopted a different definition of barriers to entry, at least for the purpose of antitrust analysis. They argue that a barrier to entry should refer to an advantage that the incumbent firm has that an entrant cannot secure.[194] The $100 million plant is not an example of such a barrier because anyone with the money can build one. The incumbent does not have any special advantage. It may be that the addition of a third plant would drive price so low that it is not worth building a plant. That, however, is an indicator that it is not socially worthwhile to use society's scarce resources to build another plant.[195]

Which definition is more appropriate? For most antitrust issues we are interested in knowing whether incumbent firms can earn supra-competitive returns without attracting entry. From society's standpoint we would like to attract entry into various lines of business until the risk-adjusted return from entry is competed down to the competitive level. The fact that significant investments are required to enter an industry does not provide any information on this issue. Incumbents could be earning a risk-adjusted competitive rate of return in an industry as a result of having made significant risky investments over time. The fact that it is costly for another firm to enter the industry is not evidence of a problem. If demand expanded unexpectedly incumbents would temporarily earn supra-competitive returns as a result of raising prices. Either they could expand capacity by making more investments or an entrant could do so. With today's highly liquid capital markets there is no basis for arguing that firms cannot secure the funding for investments that will yield a more than competitive rate of return. Consequently, the first definition—an entry barrier means it is "hard" to enter—does not advance the ball.

Suppose, however, that entrants lack some special advantages that the incumbents had. In this case the incumbents would be earning a return from those special advantages. For example, perhaps the incumbent has a patent that raises the cost of entry—maybe the entrant has to invest in a costly design-around; maybe the entrant cannot produce as good a product; or maybe the entrant will have higher costs of production. The patent is not necessarily a barrier to entry under the second definition even though it may enable the incumbent to earn what seems, *ex post*, to be a supra-competitive rate of return. It is possible that the entrant could obtain similar advantages for itself by making a similar risky investment. But the patent may be a barrier—suppose the patent is not really the result of risky investment but simply the result of a sloppy (but costly to overturn) decision by the U.S. Patent Office.

To take another example, suppose the incumbent received tax breaks that enabled it to build its plant for $50 million. If the entrant does not receive those same tax breaks it could never enter the industry on the same cost terms as the incumbent. In effect, the incumbent would end up receiv-

[194] *See* George Stigler, The Organization of Industry (1968). *See also* Carlton & Perloff, *supra* note 6, at 76-77. Unlike the Bain definition one does not need to inquire into whether the "barrier" gives rise to a supra-competitive rate of return. One only has to ask whether the alleged barrier is something that entrants could obtain at the same cost as the incumbent.

[195] For a survey of court decisions pertaining to barriers to entry, see Werden, *supra* note 194.

ing a return on its tax breaks. Whether one considers that return to be supra-competitive would depend on the details of how the tax breaks came to be. A related example involves government policies or regulations that make it harder for entrants to come in. For many years the Federal Communications Commission (FCC) made it extremely difficult for firms to compete in long-distance telecommunications. Only legal challenges forced the FCC ultimately to permit the entry of MCI.[196] The FCC prohibitions were a barrier to entry under the second definition.

U.S. v. Microsoft provides an interesting example since the case involved a two-sided market (operating systems) and barriers to entry figured prominently.[197] The government argued that the stock of applications that had been written for Microsoft Windows was a barrier to entry that prevented other firms from competing in the market for operating systems for Intel-compatible computers.[198] Applications developers write to Windows because a large number of end-users, and end-users use Windows because it has a large number of applications. The district court agreed with the government.[199]

This analysis of barriers to entry is not controversial under the first definition. Developing an operating system that is attractive to end-users and software developers is expensive.[200] However, as mentioned above, the fact that this development is expensive does not really provide any useful information. First, expensive is a relative concept. The costs of entering the two-sided market for operating systems would need to be weighed against the profits that could be earned in this market. Microsoft's annual revenues from its desktop operating system products exceeded $8 billion in 2001.[201] Second, the fact that it is expensive to enter does not provide any information on whether entrants are deterred from making investments that could reduce any supra-competitive returns that Microsoft realizes.

[196] Gerald Brock, Telecommunication Policy for the Information Age: From Monopoly to Competition (1994).

[197] *See* United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001).

[198] United States v. Microsoft Corp., Civil Action Nos. 98-1232 and 98-1233 (TPJ), Findings of Fact, Nov. 5, 1999, at ¶ 36.

[199] United States v. Microsoft Corp., Civil Action Nos. 98-1232 and 98-1233 (TPJ), Direct Testimony of Franklin M. Fisher, Oct. 14, 1998, at ¶ 70.

[200] For instance, it took about three years for Linus Torvalds and the open source programmers contributing to Linux to develop version 1.0 of the Linux Kernel. See What is Linux, Linux Online (visited Mar. 22, 2002) <http://www.linux.org/info/index.html>. It took more than four years for Microsoft to release version 1.0 of Microsoft Windows. *See* Windows Timeline (visited Aug. 15, 2002) <http://www.technicalminded.com/windows_timeline.htm>. It took more than two years of mutual effort by Microsoft and IBM to develop OS/2. *See* OS/2 1.0 (visited Aug. 15, 2002) <http://pages.prodigy.net/michaln/history/os210/index.html>. It took approximately four years to develop LisaDesk, the first Apple operating system based on Alto research of Xerox Corporation. See History of Apple Macintosh Operating System, (visited Aug. 15, 2002) <http://perso.club-internet.fr/jctrotot/Perso/History.html>.

[201] Microsoft Corp. 10-K (FY2001) (downloaded Aug. 30, 2002) <http://www.sec.gov/Archives/edgar/data/789019/000103221001501099/0001032210-01-501099.txt>.

The government's testimony and the district court's opinion suggest that they were analyzing barriers under the more stringent but more relevant second definition.[202] They argued that Microsoft had special advantages over rivals because it was the first entrant into the market for operating systems for Intel-compatible computers.[203] Applications developers wrote for Microsoft's operating system because it was the only alternative. Microsoft did not have to expend resources to persuade application developers to write for its operating system versus other alternatives.

Let us consider this analysis through the lens of two-sided markets. Participation in a two-sided market naturally requires the development of customers on both sides. The fact that Microsoft has attracted a large number of applications is no more or less remarkable than the fact that it has attracted a larger number of users. It is also no more remarkable than the fact that Sony has attracted a large number of game developers, and that ATM networks have a large number of ATM machines that connect with their networks throughout the country.

Indeed, the fact that a firm in a two-sided market has developed customers on one side is, from society's standpoint, a good thing. Not only do these customers benefit directly from the services they obtain—they are inputs into the provision of services to customers on the other side. The fact that Microsoft has a large stock of applications reflects the fact that applications developers are receiving useful services from the features included in Microsoft's operating system. It also reflects considerable value to end-users who, as the government and district court correctly observed, value these applications. There is no basis for suggesting that it is a bad thing for Microsoft to invest resources in getting more applications written for Windows, anymore than it is a bad thing for Nintendo to try to get more games written for its console, for Discover to try to get more merchants to take its card, or for Goldman Sachs to try to find more companies that want to go on the block.

By themselves positive feedback effects do not prevent entry. Indeed, many two-sided markets have multiple providers. In the case of payment cards, there was successive entry by Diners Club (1950), American Express (1958), Visa (1966), MasterCard (1966), and Discover (1986).[204] In the case of video games in the US, there was successive entry by Magnavox (1972), Atari (1975), Coleco (1976), Fairchild (1976), Mattel (1979), Nintendo (1985), Sega (1989), Sony (1995), and Microsoft (2001).[205]

It could be that Microsoft has special advantages that constitute a barrier to entry under the second definition. But neither the government nor the district court put forward an analysis that could distinguish these special advantages, if any, from the investments that Microsoft made in getting and keeping both sides on board. The relevant question to ask concerning barriers to

[202] *See* Findings of Fact, *supra* note 198, at 13-16.
[203] *Id*, at 16. ("Microsoft never confronted a highly penetrated market dominated by a single competitor").
[204] *See* Evans & Schmalensee, *supra* note 10, at 10, 62-66.
[205] *See* KENT, *supra* note 105, at xi-xvi.

entry into a two-sided market is whether the incumbent has special advantages that could not be replicated by entrants. That question needs to focus on the costs of getting into both sides of the market. Saying that Microsoft has a lot of applications written for its operating system is no more probative than saying that Dating Club B has a lot of women lined up at the door or that American Express has many companies signed up for its corporate charge cards. Getting customers on both sides is the name of the game in two-sided markets.

One way to think of this is to consider two operating system companies. Entrant 1 comes in before Entrant 2. To get both sides on board, the company with the head start, Entrant 1, has to spend $1 billion to get developers to write applications. If Entrant 2 had to spend $1.1 billion to get both sides on board we would probably conclude that the entry barrier is fairly modest relative to the risk-adjusted profits that could be earned in this business. If Entrant 2 had to spend $2 billion we might reach the opposite conclusion. In both cases we would want to consider these entry barriers relative to prospective profits.

IBM's experience in entering the operating system business with OS/2 is instructive. Shortly after the introduction of Windows 2.0, OS/2 was completed. However, due to its high price and incompatibility with other existing applications it was deemed a failure.[206] If IBM had made the same investment as Microsoft in getting both sides of the market on board it is not clear that, with an equal or superior operating system, it would not have succeeded. In any case, my point is not that there were or were not barriers to entry to the operating system market. It is that by ignoring the economics of two-sided markets the government failed to conduct the relevant inquiry.

The government also relied on the applications barrier to entry in its liability theory. It argued that Netscape and Java would help erode the applications barrier to entry by allowing software developers to write to a middleware layer that would sit between Windows and the applications.[207] The words here matter. The phrase applications barrier to entry is a pejorative term that suggests that the stock of applications served primarily as an entry barrier. The Government's chief economic witness in *U.S. v. Microsoft* has written:

> That positive feedback effect—the applications barrier to entry—has made it difficult or impossible for rival operating systems to compete effectively with Microsoft by gaining more than a niche in the market. New entry is not likely to erode Microsoft's market share and market power as long as the applications barrier to entry remains strong[.][208]

---

[206] Martin Campbell-Kelly, *Not Only Microsoft: The Maturing of the Personal Computer Software Industry, 1982-1995,* 75 BUS. HIST. REV. 103 (2001), *available in* NEXIS, at 12. *See also* United States v. Microsoft Corp., 253 F.3d 34, 55 (D.C. Cir. 2001).

[207] *See* Direct Testimony of Franklin M. Fisher, *supra* note 199, at ¶¶ 36, 95.

[208] David S. Evans *et al.*, Did Microsoft Harm Consumers? Two Opposing Views 11 (2000).

And the district court in U.S. v. Microsoft stated that:

> [T]he applications barrier would prevent an aspiring entrant into the relevant market from drawing a significant number of customers away from a dominant incumbent even if the incumbent priced its products substantially above competitive levels for a significant period of time.[209]

The theory of two-sided markets sees the stock of applications in neutral terms—it is simply the amount of demand satisfied on the developer side of the market and the value that can be brought to user side of the market. All two-sided firms want to increase demand on both sides and increase their advantages on both sides over rivals. Visa has more merchant acceptance than American Express, Sony has more videogames written for its console than Microsoft, and Time magazine has more readers than The New Republic.

D. *Per Se* vs. the Rule of Reason—the Role of Efficiencies

Efficiencies play an important role in evaluating antitrust matters. In the merger context, the social benefits of economies of scale and scope weigh against the social costs of price increases through reduced competition; these economies may be so large that consumers benefit from lower prices even after accounting for price increases from reduced competition. In cases involving a full-blown rule-of-reason analysis, the courts consider whether the efficiencies that result from challenged practices outweigh their anticompetitive effects. Finally, in cases involving practices that are usually considered *per se* illegal, the courts consider whether efficiencies are so pronounced that the practices should be analyzed under the rule of reason.

Two special issues concerning efficiencies arise in two-sided market cases. The first concerns the benefits that consumers in each group receive as a result of having access to the other group of consumers. The second concerns the benefits that consumers receive from practices that are either essential for getting both sides on board or that get both sides on board at lower costs than alternative practices. This issue arises when coordination among competitors is used to establish the pricing structure.

1. Benefits from Indirect Network Effects

The net value—benefits less price paid—that each consumer receives from purchasing a product with network effects increases with the number of other consumers that also use that product. Any analysis of antitrust issues in markets with network effects should take these benefits into account. In two-sided markets, the net value increases with the number of consumers on the other side of the market.

The value of indirect network effects are likely to be substantial in many contexts. That is because each consumer is providing a benefit (directly in the case of matching markets like dat-

[209] *See* Findings of Fact, *supra* note 198, at ¶ 36.

ing clubs, indirectly in the case of other markets such as operating systems). Some numerical examples can demonstrate the possible magnitudes of the effects. Using the Rochet-Tirole formulation above, consider a two-sided market with 100 "buyers" and 100 "sellers," which would lead to 10,000 (100 × 100) transactions. Let us also assume that buyers (and sellers) have preferences such that the 100th (or marginal) buyer has a 1 cent benefit per transaction, the 99th has a 2 cent benefit, and so on, up to the 1st buyer, which has a $1.00 benefit per transaction. It is straightforward to show that the per-transaction benefit received by the average buyer is 50.5 cents.

The 100th buyer gets relatively little benefit from the 100 transactions it engages in, only 100 × 1 cent = $1.00. On the other hand, its decision to join has benefited each of the 100 sellers. The 100th seller only benefits as much as the 100th buyer for that transaction, but everyone else benefits by much more. The 100 transactions generated by the last buyer produces benefits on the seller side averaging 50.5 cents per transaction, for a total benefit of $5.05. That is, the indirect benefits to sellers are more than 5 times the direct ones to the 100th buyer. On average, across all buyers and sellers, the indirect benefit is essentially equal to the direct benefit.[210]

The merger of two firms in a two-sided market is an obvious place in which competition regulators should consider the efficiencies from the merger as well as its prospect for increasing prices. ATM network mergers serve as a good example.[211] In rule of reason cases the courts need to examine the effect of the challenged practice on consumer demand on each side of the market and the indirect effects from each side to the other side. Consider the *Visa Check/MasterMoney* litigation.[212] Plaintiffs claim that Visa and MasterCard have policies that result in a tie between credit cards and debit cards and that without this tie the interchange fee to merchants would be lower. Let us suppose this is correct. That cannot be the end of the analysis because one must analyze the effect of the practice on the inextricably linked sides of the market. A forced reduction in the interchange fee to merchants would reduce the stream of revenues to banks that issue debit cards; under competition, these banks would increase the fees they charge for debit cards; that it turn would reduce the number of debit cards held and used; and that in turn would reduce the value that merchants get from debit cards.[213]

---

[210] The Rochet-Tirole model assumes buyers are not more likely to join if there are more sellers (or vice versa). If we assumed the base of sellers were important to attracting buyers (and vice versa), the indirect benefits would be even greater because a buyer joining the system would induce additional sellers to join (and so on), which would generate additional indirect benefits on both the seller and buyer sides.

[211] Robin A. Prager, *ATM Network Mergers and the Creation of Market Power*, ANTITRUST BULL. 355 (Summer 1999). Some examples of ATM mergers include Peak-Minibank (1996), Exchange-Instant Teller (1996), and SCS-Pacific Interchange (1992). See *supra*, at 361.

[212] *See In re* Visa Check/MasterMoney Antitrust Litig.; Wal-Mart, *et al*. v. Visa U.S.A. Inc. and MasterCard International, Inc., 280 F.3d 124 (2d Cir. 2001)

[213] Based on the arguments presented at class certification the plaintiffs appear to be arguing that it is possible to change one element of the pricing structure without having significant effects on the other elements. They argued that "under the particular circumstances of the market at issue in this case, credit card interchange fees would not have increased in the 'but-for,' untied world." See In re Visa Check/MasterMoney Antitrust Litig.; Wal-Mart, *et al*. v. Visa U.S.A. Inc. and MasterCard International, Inc., 280 F.3d 124, 154 (2d Cir. 2001).

2. Methods for Getting both Sides on Board

Some two-sided markets have competing sellers on both sides of the market but have developed institutions that provide the coordination necessary to get both sides on board. One example is residential real estate. Multiple Listing Services (MLS) provide a central location in which selling agents can post properties and buyer agencies can locate these properties. MLS agencies and members are able to show all of the listed homes to the buyer side of their respective client bases.[214] Thus, retaining one MLS realtor enables the sellers to attract a larger number of offers than would otherwise be possible through retaining each of the individual member agencies.[215] Another example is payment cards. There are many associations around the world of entities that issue payment cards or service merchants.[216] The association typically establishes an interchange fee that ultimately determines the price structure—the relative prices to cardholders and merchants. An alternative to these institutions would be to have a single firm provide the coordination as well as supplying both sides. Some of the payment-card firms were organized just this way—Diners Club, American Express, and Discover choose both pricing levels and pricing structures; MasterCard and Visa, as associations, choose pricing structures but let the pricing levels emerge through competition among their members.

Coordination among competitors over prices is usually a *per se* violation of the antitrust laws. The courts have recognized that it is more appropriate to analyze a practice under the rule of reason if there is reason to believe that significant efficiencies may result from this practice.[217] Two issues arise in analyzing whether there are significant efficiencies through coordinated pricing in two-sided markets. The first is whether there are absolute efficiencies from the practice: does the coordinated pricing in fact provide a mechanism for getting both sides on board and for choosing an optimal pricing structure? Agreements on pricing levels as opposed to pricing structures will often fail this test. The second is whether there are relative efficiencies from the practice: is a market structure with coordinated pricing more efficient than a market structure without coordinated pricing? For example, MasterCard could operate as a single firm rather than an association—it would then set the price level and the price structure. We suspect in most cases agreements that are absolutely efficient are also relatively efficient. That is because moving from coordination among competitors to non-coordination over pricing structures will generally decrease the number of competitors that determine the price level.

---

[214] Frew and Jud, *supra* note 33, at 178.

[215] *Id.*

[216] Many countries have associations that serve those countries but belong to one of the global associations for the purpose of having their cards accepted outside the country. For example, banks belonging to Groupement des Cartes Bancaires in France issue the "CB" card. There is a domestic version only on Cartes Bancaires' domestic network, but most cards have Visa or MasterCard logos and can be used at merchants outside of France that accept Visa or MasterCard. See Cartes Bancaires website (visited Sep. 2, 2002) <http://www.cartes-bancaires.com/GB/Pages/FrameVie.htm>.

[217] *See* Broadcast Music, Inc. v. CBS, 441 U.S. 1 (1979).

3. Antitrust Analysis of Payment-Card Interchange Fees

The role of coordinated pricing in two-sided card markets has been considered by the U.S. courts in *NaBanco v. Visa U.S.A.*[218] In a case that was filed in 1979, National Bancard Corporation (NaBanco) claimed that the interchange fee was a price fixing agreement that violated of Section 1 of the Sherman Act. Visa did not deny that it had fixed prices. However, argued that setting an interchange fee was necessary for the existence of the product sold by the association.[219] Visa explained the role of the interchange fee in what we would now call a two-sided market.[220] It argued that the purpose of the interchange fee was to provide a mechanism "to distribute the costs of the system in relation to prospective benefits so as to encourage members to engage in the appropriate balance of card-issuing and merchant-servicing."[221]

The district court and the Eleventh Circuit agreed with Visa. The Eleventh Circuit upheld the district court's rule-of-reason approach in the case, explicitly relying on the two-sided nature of the industry:

> Another justification for evaluating the IRF under the rule of reason is because it is a potentially efficiency creating agreement among members of a joint enterprise. There are two possible sources of revenue in the VISA system: the cardholders and the merchants. As a practical matter, the card-issuing and merchant-signing members have a mutually dependent relationship. If the revenue produced by the cardholders is insufficient to cover the card-issuers' costs, the service will be cut back or eliminated. The result would be a decline in card use and a concomitant reduction in merchant-signing banks' revenues. In short, the cardholder cannot use his card unless the merchant accepts it and the merchant cannot accept the card unless the cardholder uses one. Hence, the IRF accompanies "the coordination of other productive or distributive efforts of the parties" that is "capable of increasing the integration's efficiency and no broader than required for that purpose."[222]

The Eleventh Circuit went on to find that "[a]n abundance of evidence was submitted from which the district court plausibly and logically could conclude that the IRF on balance is procompetitive because it was necessary to achieve stability and thus ensure the one element vital to the survival of the VISA system—universality of acceptance."[223] The Supreme Court declined to review the Eleventh Circuit's decision. Although the economic analysis of two-sided markets cannot

[218] National Bankcard Corp. (NaBanco) v. VISA U.S.A., Inc., 779 F.2d 592 (11th Cir. 1986).
[219] It relied on Broadcast Music, Inc. v. CBS, 441 U.S. 1 (1979).
[220] One of Visa's consultants went on to write one of the earliest papers on the economics of two-sided markets. *See* William F. Baxter, *Bank Interchange of Transactional Paper: Legal and Economic Perspectives*, 26 J.L. Econ. 541 (1983).
[221] National Bancard Corp. (NaBanco) v. VISA U.S.A., Inc., 596 F. Supp. 1231, 1260 (S.D. Fla. 1984).
[222] National Bankcard Corp. (NaBanco) v. VISA U.S.A., Inc., 779 F.2d 592, 602 (11th Cir. 1986).
[223] *Id* at 605.

by itself say whether Visa's interchange fee resulted in the optimal pricing structure from society's standpoint, the Eleventh Circuit correctly recognized the economic role of interchange fees in devising a price structure that took both sides of the market into account.

The Royal Bank of Australia (RBA) reached a different conclusion in a recent investigation.[224] It was concerned that Visa's interchange fees were set too high, so as to encourage overuse of credit cards in Australia. It relied on the findings in economic models of interchange fees that firms may have private incentives to set the fees higher than socially optimal.[225] The reasoning is that an individual merchant may have an incentive to accept Visa cards if it expects to make sufficient incremental sales from doing so, even if Visa cards are more expensive for the merchant than its other alternatives. The RBA argued that these incremental sales were not a social benefit, as they came at the expense of other merchants. For all merchants collectively, acceptance of Visa cards may have generated little or no incremental sales. Thus, the argument was that Visa could exploit each individual merchant's willingness to pay, which derived from private not social benefits. In response to these concerns, the RBA proposed a regulatory scheme for interchange fees that was based on an allocation of cost to the issuer and acquirer sides; the scheme does not consider demand.

The RBA had not established that interchange fees were too high. It relied on the existence of theoretical models that showed that they could be too high. But the same models also showed that privately set interchange fees could be at the socially optimal level or lower. It makes no sense to seek to lower interchange fees when we do not know if they are too high. All we know, and this should come as no surprise in any two-sided market, is that one side (the merchants in this case) would prefer to pay less. The RBA's proposed regulatory scheme has no basis in the economics of two-sided markets. The socially optimal price structure depends on a complex contribution of costs and demand.[226] There is no basis for focusing only on costs, nor is there a basis for assigning costs to one side or the other. There is no economic basis for believing that the RBA method for determining the interchange fee would increase or decrease the overall welfare of the consumers in the two sides of the market.

E. Predation

The recognition that business strategies and their effects on consumers must be evaluated with respect to both sides of the market has important implications for the analysis of predation. It may be privately and socially optimal for prices on one side of the market to be below any possible measure

[224] *Reform of Credit Card Schemes in Australia IV*, Finals Reforms and Regulation Impact Statement, Reserve Bank of Australia, Aug. 2002.

[225] *Reform of Credit Card Schemes in Australia I: A Consultation Document*, Reserve Bank of Australia, Dec. 2001, at section 2.4.

[226] *See generally* Ahlborn *et al., supra* note 21; Schmalensee (2002), *supra* note 4; Rochet & Tirole, *supra* note 4.

of cost on this side. That is true not only during the initial stage in which most economists and courts have recognized the virtues of "penetration pricing."[227] It is also true during the long-run equilibrium of the industry. Complaints by entrants that they cannot get into one side because the incumbent is engaging in predatory pricing—or has made predatory investments—are likely to have little merit in two-sided markets. It also may be privately and socially optimal for firms to make significant investments in one side even though these investments do not appear to generate profits on that side. Again, this can occur even when the firm is mature.

To clarify the issues, let us consider extending the Brooke Group test of predatory pricing to two-sided markets.[228] The test has two prongs. First, a plaintiff alleging predation must show that the defendant's prices were "below an appropriate measure of… costs."[229] Thus, pricing must be below cost to support a claim of predation, even though, in theory, there can be predatory prices that are above cost. Second, the plaintiff must show that the defendant had "a reasonable prospect, or, under §2 of the Sherman Act, a dangerous probability, of recouping its investment in below-cost prices."[230] Finding that a competitor was harmed—even driven from the market—is not enough under this test. Moreover, for a dangerous probability of recoupment to exist, low prices must eliminate substantial competition in a way that persists even after a post-predation price increase.

For two-sided markets the pricing prong of this test should be based on the defendant's pricing level—the overall price to the two markets. This is simple and straightforward when both sides of the market have the same unit of observation (person, transaction, product, etc.). One would calculate whether the total price charged for a given transaction, let us say, exceeds marginal cost (or, following the Areeda-Turner analysis, average variable cost).[231]

This analysis would apply to payment cards, dating services, B2B exchanges, real estate and other matching-type markets. When the two sides of the market have different units of observation one could normalize them to the same unit by basing the calculation on total revenue and total variable cost. If total variable cost is less than total revenue then the defendant would fail the first prong of the Areeda-Turner version of the Brooke Group test.

The recoupment prong is straightforward. For two-sided markets the court would need to consider whether there is a dangerous probability that the defendant will raise prices high enough and for long enough on both sides of the market to recoup its losses during the alleged predatory phase. There is nothing novel about this for two-sided markets other than accounting for the two sides.

---

[227] *See*, e.g., Brooke Group Ltd. v. Brown & Williamson Tobacco Corp. 509 U.S. 209 (1993); Phillip Areeda & Donald F. Turner, *Predatory Pricing and Related Practices under Section 2 of the Sherman Act*, 88 HAR. L. REV. 697 (1975).
[228] *See* Brooke Group Ltd. v. Brown & Williamson Tobacco Corp. 509 U.S. 209 (1993).
[229] *Id*. at 222.
[230] *Id*. at 224.
[231] *See* Areeda & Turner, *supra* note 227, at 702, 716-18.

An example of predation in a two-sided market was suggested by the government in *U.S. v. Microsoft*. Microsoft gave away a software tool to make it easier for Internet Service Providers (ISPs) and corporate IT departments to customize Internet Explorer. Netscape sold a similar kit for $1,995.[232] Microsoft had a two-sided justification: it gave away the tool to increase ISP demand (one side of the market) and provide other features to increase the demand for Windows (which included Internet Explorer) among users. The Appeals Court rejected the government's predation claim here, although mainly because of its general skepticism about low prices being anticompetitive.[233] The government made several other claims concerning predatory pricing and investments—all these claims ignored the two-sided nature of the market.[234]

F. Other Antitrust Issues

Business practices and relationships in two-sided markets tend to be more complex than in one-sided markets. First, firms are more likely to operate as multi-product firms with feet in both markets. More generally, because the production of complementary products is important in two-sided markets, firms may end up producing a variety of other products that stimulate demand on one side of the market or the other. For example, many producers of operating systems also produce application programs for end-users in addition to providing software interfaces for external software developers to use; many payment-card companies own companies that provide inputs for the production of services to cardholders or merchants; and several video game companies develop games to run on their consoles.

Second, sometimes there is coordination among competitors—especially when there is what we have referred to as multihoming. One or both sides of the market can benefit when there is a standard technology or protocol that enables them to use products from multiple vendors. Two-sided firms have conflicting profit incentives—they would like to discourage standardization to increase their own market power; they would like to encourage standardization if this will expand overall demand sufficiently. American Express and Visa are both members, for example, of Global Platform, an international organization that sets standards for smart card technology,[235] and are using Global Platform standards in their respective smart card development efforts.[236]

---

[232] *See* Findings of Fact, *supra* note 198, at ¶ 250.

[233] "The rare case of price predation aside, the antitrust laws do not condemn even a monopolist for offering its product at an attractive price, and we therefore have no warrant to condemn Microsoft for offering either IE or the IEAK free of charge or even at a negative price. Likewise, as we said above, a monopolist does not violate the Sherman Act simply by developing an attractive product." United States v. Microsoft Corp., 253 F.3d 34, 68 (D.C. Cir. 2001).

[234] United States v. Microsoft Corp., Civil Action No. 98-1232, Complaint, May 18, 1998, at ¶ 16.

[235] *See* the Global Platform website (visited Aug. 15, 2002) <http://www.globalplatform.org/news2.asp>.

[236] *See* Global Platform Specification (visited Aug. 15, 2002) http://international.visa.com/fb/paytech/productsplatforms/globalplatform.jsp>; and the Global Platform Brochure (downloaded Aug. 15, 2002) <http://www.globalplatform.org/pdf/GlobalPlatform_Brochure.pdf>.

Third, two-sided firms sometimes take actions to coordinate the behavior of their customers. That occurs because standardization by one set of customers benefits the other set of customers. For instance, recently B2Bs have been moving towards the standardization of information that might significantly enhance and automate various procedures such as RFPs (request for proposals), RFQs (request for quotes), fax requests, phone inquiries and purchase orders.[237] NMMs, third-party intermediaries whose primary purpose is to match corporate buyers and sellers, play a pivotal role in this process.[238] In *U.S. v. Microsoft*, Microsoft objected to letting computer manufacturers, or for that matter end-users, selectively delete portions of the operating system code. One of its justifications was that applications developers would then lack assurance on what features are available for them to rely on in the operating system on an end-user's machine.[239]

Fourth, associations of competitors, joint ventures, and standard-setting bodies may arise to provide coordination among competitors or among customers or among other important input suppliers to either side of the market. Payment-card associations operate the network and set rules that result in the determination of a pricing structure. Real estate agencies have associations that operate the MLS.[240] Microsoft collaborates with Intel and many hardware manufacturers to help ensure that the Wintel platform attracts users and developers.[241]

These practices tend to be flypaper for antitrust complaints. The multi-product nature of firms in two-sided markets can lead to a variety of issues surrounding "vertical relationships." For example, a firm might be accused of trying to dominate one side of a two-sided market to reduce competition on the other side. Or to take another example, suppose a firm on one side of a two-sided market wants to deliver customers who possess products A and B to the other side. One way it might do that is to integrate A and B or require that customers purchase B as a condition of purchasing A. The firm might be accused of tying in this instance—this is the plaintiff's claim in the *Visa Check/MasterMoney*.[242] Although these sorts of vertical relationships may well result in anti-competitive effects, it is necessary to evaluate their effects in the context of competition in two-sided markets.

---

[237] Roopen Roy & Abhijit Roy, *Net Market Makers: The Winners in the B2B Play!* Indianinitiative.com, Dec. 21, 2001, (visited Aug. 21, 2002) <http://www.indiainitiative.com/indiainitiative_new/e-business/b2b_abhi_roopen.htm>.

[238] *Id*.

[239] Ronald A. Cass, *Copyright, Licensing, and the 'First Screen,'* 5 Mich. Telecommunications Tech. L. Rev. 35, 59-64 (1999).

[240] Real estate boards are non-profit organizations representing local real estate agents/brokers, which operate Multiple Listing Services in local communities. For a definition, see Home and Real Estate Glossary (visited Aug. 28, 2002) <at http://www.homes-and-real-estate.com/glossary/r.htm>.

[241] Carmen Nobel, *Microsoft, Intel Seek to Become Wireless Force*, eWeek, Feb. 19, 2002 (visited Aug. 28, 2002) <http://www.eweek.com/article2/0,3959,108489,00.asp>.

[242] *In re* Visa Check/MasterMoney Antitrust Litig.; Wal-Mart, *et al.* v. Visa U.S.A. Inc. and MasterCard International, Inc., 280 F.3d 124 (2d Cir. 2001).

Coordinated behavior also raises antitrust concerns. Whenever businesses get together there is a worry that they are colluding to raise price or otherwise harm consumers. That may be just as true in two-sided markets as it was for Adam Smith's tradesmen.[243] Nevertheless, there are legitimate pro-competitive reasons for coordination among competitors or among customers and for having institutions to help do this in two-sided markets.

V. CONCLUSIONS

Two-sided markets are an increasingly important part of the global economy. Firms that provide platforms for multiple customer groups are a critical part of many interrelated segments of the computer industry. These "keystone" firms include Microsoft in operating systems for Intel-compatible computers and Intel for computer chips. In most industrialized countries a large fraction of payments takes place through firms and associations that provide platforms for merchants and customers to exchange money. The increased importance of the Internet for household-to-household, business-to-household, and business-to-business transactions and the emergence of e-pay systems on the Internet will increase the fraction of payments going through commercial payment platforms. And speaking of the Internet, although dot-coms are currently in a slump, Internet-based businesses are sure to flourish over time and many of these are likely to be based on a two-sided model.

But two-sided markets are not just in these high-profile sectors. They are dotted across the economy. We began with perhaps a trivial example of dating clubs—discotheques, church clubs for singles, and local village matchmakers could have served just as well. Others range from real estate to video games to media firms. Some of the most recognizable brands in the world operate in two-sided markets: Bloomberg, Century 21, Microsoft, Sony and Visa.

Two-sided markets and firms in those markets behave in ways that seem surprising from the vantage point of traditional industries, but in ways that seem like plain common sense once one understands the business problems they must solve. "Getting both sides on board." "The chicken-and-egg problem." These are the mantras ones hears from the entrepreneurs in these industries, the trade press that covers them, business gurus, and journalists. They contain important economic implications.

Two-sided firms have to come up with the right price structure and the right investment strategy for balancing the demands of the customer groups they must get and keep on their platforms. That is a different problem than is faced by one-sided firms. It is also harder. One way to see that is to recognize that different firms have chosen different price structures and have realized different fortunes from their choices. American Express bet on a price structure skewed against

[243] Adam Smith, an Inquiry into the Nature And Causes of the Wealth of Nations ¶ I.10.82 (Edwin Cannan ed., 1904).

merchants—it worked for some years but then got it into trouble.[244] Meanwhile Visa surpassed American Express, a firm that was once dominant and seemed unbeatable.[245] Microsoft bet on a price structure that catered to software developers. Bloomberg bet on a simple formula for its data terminals—a flat fee for subscribers, and few charges for content providers.

There is no reason to believe that anticompetitive problems are more prevalent or less prevalent in two-sided industries than other ones. Likewise, there is no basis for asking regulators to steer clear of these industries or to spend extra effort on them. The message of this chapter, however, is that antitrust analysis of these industries should heed the economic principles that govern pricing and investment decisions in these industries. Prices do not and prices cannot follow marginal costs in each side of the market. Price levels, price structures, and investment strategies must optimize output by harvesting the indirect network effects available on both sides. By doing so businesses in two-sided industries get both sides on board and they solve the chicken-and-egg problem.

[244] Between mid-1980s and 1996 the American Express charge and credit card share dropped from more than 24 percent to 16 percent. By mid-1990s American Express realized the necessity to adopt a new business model. *See* Evans & Schmalensee, *supra* note 10, at 185-193.

[245] By 1996 Visa charge and credit card share was more than 45 percent compared to 16 percent held by American Express. *Id*. at 174, 187.

APPENDIX A

TABLE 1 Sources of Platform Revenue in Selected Two-Sided Platforms

Industry	Two-Sided Platform	Side One	Side Two	Side That Gets Charged Little	Sources of Revenue
Real Estate	Residential Property Brokerage	Buyer	Seller	Side One	Real estate brokers derive income principally from sales commissions.[1]
Real Estate	Apartment Brokerage	Renter	Owner/ Landlord	Typically Side One	Apartment consultants and locater services generally receive all of their revenue from the apartment lessors once they have successfully found tenants for the landlord.[2]
Media	Newspapers and Magazines	Reader	Advertiser	Side One	Approximately 80 percent of newspaper revenue comes from advertisers.[3]
Media	Network Television	Viewer	Advertiser	Side One	For example, FOX earns half of its revenues from advertisers.[4]
Media	Portals and Web Pages	Web "Surfer"	Advertiser	Side One	The average portal gets slightly over half of its revenues from advertisements. All other web pages generally receive about a tenth of their revenue from advertisements.[5]
Software	Operating System	Application User	Application Developer	Side Two	For example, Microsoft earns at least 67 percent of its revenues from licensing packaged software to end-users.[6]
Software	Video Game Console	Game Player	Game Developer	Neither – Both sides are significant sources of platform revenue.	Both game sales to end users and licensing to third party developers are significant sources of revenue for console manufacturers.[7]
Payment Card System	Credit Card	Cardholder	Merchant	Side One	For example, in 2001, American Express earned 82 percent of its revenues from merchants.[8]

Notes: [1]*Real Estate Brokers and Sales Agents, in* Occupational Outlook Handbook, U.S. Department of Labor, Bureau of Labor Statistics (visited Sept. 3, 2002), <http://www.bls.gov/oco/ocos120.htm>; [2]Courtney Ronan, *Apartment Locaters: How Do They Make Their Money?* Realty Times website, Jun. 30, 1998 (visited Aug. 22, 2002) <http://realtytimes.com/rtnews/rtcpages/19980630_aptlocator.htm>; [3]Lisa George and Joel Waldfogel, Whom Benefits Whom in Daily Newspaper Markets? NBER WORKING PAPER #7944 11 (October 2000); [4]Brian O'Keefe, *Extra! Extra! Hot Media Stocks!* Fortune, June 24, 2002 (visited Aug. 22, 2002) <http://www.pbs.org/wsw/news/fortunearticle_20020620_01.html>; [5]Clicking into Orbit, Technology Trend Quarterly Report, March 2000 (visited Aug. 15, 2002) <www.deloitte.nl/downloads/website/clicking_into_orbit.pdf>; [6]*1994 Worldwide Software Review and Forecast,* IDC Report #9358, November 1994, *1995 Worldwide Software Review and Forecast,* IDC Report # 10460, November 1995, *1996 Worldwide Software Review and Forecast*, IDC Report #12408, November 1996, *1997 Worldwide Software Review and Forecast*, IDC Report #14327, October 1997, *1999 Worldwide Software Review and Forecast*, IDC Report #20161, October 1999, *Worldwide Software Market Forecast Summary, 2001-2005*, IDC Report #25569, September 2001 ; [7]David Becker, *Revenue from Game Consoles will Plunge, Report Predicts*, CNET.Com, Jul. 31, 2000 (visited Aug. 5, 2002) <http://techrepublic-cnet.com.com/2100-1040-243841.html?legacy=cnet>; [8]American Express Company Annual Report 2001 (visited Aug. 15, 2002), <http://www.onlineproxy.com/amex/2002/ar/pdf/axp_ar_2001.pdf>, at 35.

TABLE 2 The Presence of Multihoming in Selected Two-Sided Platforms

Two-Sided Platform	Side One	Presence of Multihoming for Side One	Side Two	Presence of Multihoming for Side Two
Residential Property Brokerage	Buyer	*Uncommon:* Multihoming may be unnecessary, since an MLS allows buyers to see property listed by all member agencies.[1]	Seller	*Uncommon:* Multihoming may be unnecessary, since an MLS allows the listed property to be seen by all member agencies' customers.[1]
Securities Brokerage	Buyer	*Common:* The average securities brokerage client has accounts at three firms.[2] Note that clients can be either or both buyers or sellers.	Seller	*Common:* The average securities brokerage client has accounts at three firms.[2] As mentioned, clients can be either or both buyers or sellers.
B2B	Buyer	*Varies:* For example, multihoming may be unnecessary for some online B2B sites, since buyers can go directly to the B2B platform instead of contacting multiple individual suppliers.[3]	Seller	*Varies:* Multihoming may be unnecessary since the B2B can inexpensively reach a large audience.[4]
P2P	Buyer	*Varies:* Multihoming may be unnecessary for buyers using online auction sites since eBay holds 85% of the market share (i.e. it seems that most people purchase their online auction products at eBay). Alternatively, multihoming may be more common for online dating services where there are many sites and a large audience of online singles at each site (considered to be available singles, as opposed to buyers).[5]	Seller	*Varies* – Multihoming may be unnecessary for sellers using online auction sites since eBay holds 85% of the market share (i.e. it seems that most people auction their products at eBay). Alternatively, multihoming may be more common for online dating services where there are many sites and a large audience of online singles at each site (considered to be available singles, as opposed to sellers).[5]
Newspapers and Magazines	Reader	*Common:* In 1996, the average number of magazines issues read per person per month was 12.3.[6]	Advertiser	*Common:* For example, Sprint advertised in the New York Times, Wall Street Journal, and Chicago Tribune, among many other newspapers, on Aug. 20, 2002.[7]

Network Television	Viewer	*Common:* For example, Boston, Chicago, Los Angeles, and Houston, among other major metropolitan areas, have access to at least four main network television channels: ABC, CBS, FOX, and NBC.[8]	Advertiser	*Common:* For example, Sprint places television advertisements on ABC, CBS, FOX, and NBC.[9]
Operating System	Application User	*Uncommon:* It is unlikely that an individual will switch to a new operating system.[10]	Application Developer	*Common:* As noted earlier, the proportions of developers that develop for various operating systems total to 205 percent, indicating significant multihoming by developers.[11]
Video Game Console	Game Player	*Varies:* The average household (that owns at least one console) owns 1.4 consoles.[12]	Game Developer	*Common:* For example, Electronic Arts, a game developer, develops for Nintendo's GameCube, Microsoft's Xbox, and Sony's Playstation 2, among other consoles.[13]
Payment Card	Cardholder	*Common:* Most American Express cardholders also carry at least one Visa or MasterCard.[14]	Merchant	*Common*: American Express cardholders can use Visa and MasterCard at almost all places that take American Express.[14]

Notes: [1]James R. Frew and G. Donald Jud, Who Pays the Real Estate Broker's Commission? in RESEARCH IN LAW AND ECONOMICS: THE ECONOMICS OF URBAN PROPERTY RIGHTS 177, 178 (Austin J. Jaffe ed. 1987); [2]Susan Scherreik, *Is Your Broker Leaving You Out in the Cold?,* Business Week Online, February 18, 2002 (visited Aug. 22, 2002) <http://www.businessweek.com/magazine/content/02_07/b3770110.htm>; [3]David Lucking-Reilly and Daniel F. Spulber, Business-to-Business Electronic Commerce, 15 J. ECON. PERSPECTIVES 55, 58 (2001); [4]*Entering the 21st Century: Competition Policy in the World of B2B Electronic Marketplaces, Part 2: Efficiencies of B2B Electronic Marketplaces,* FTC STAFF REPORT (Oct. 2000) (downloaded Aug. 14, 2002) <http://www.ftc.gov/os/2000/10/index.htm#26>, at 5; [5]Oscar S. Cisneros, *Ebay Accused of Monopolization, WIRED NEWS*, July 31, 2000 (visited Aug. 20, 2002) <http://www.wired.com/news/print/0,1294,37871,00.html> and Paul Festa, *Looking for Love Online*, CNET NEWS.COM, December 17, 1996 (visited Aug. 22, 2002) <http://news.com.com/2100-1023-255523.html?tag=rn>; [6]*Magazines in the Information Age,* FCB Media Research Report (Spring 1998), <http://www.magazine.org/resources/research/fcb_magazines_infoage.html>; [7]New York Times at A22, Wall Street Journal at A15, and Chicago Tribune at 11(sect.1); [8]ABC website (visited Sept. 3, 2002) <http://abc.abcnews.go.com/site/localstations.html>, CBS website (visited Sept. 3, 2002) <http://www.cbs.com/info/hdtv/>, FOX website (visited Sept. 3, 2002) <http://www.fox.com/links/affiliates.htm>, and NBC website (visited Sept. 3, 2002) <http://www.nbc.com/nbc/header/Local_Stations/>; [9]James Fisher, *An Ad Blitz for the 21st Century*, Sprint Press Release, September 24, 1997 (visited Aug. 23, 2002) <http://www3.sprint.com/PR/CDA/PR_CDA_Press_Releases_Detail/0,3245,,00.html?ID=1294>;[10]Scot Hacker, *He Who Controls the Bootloader*, BYTE.COM (visited Aug. 20, 2002) <http://www.byte.com/documents/s=1115/byt20010824s0001/0827_hacker.html>; [11]Josh Lerner, *Did Microsoft Deter Software Innovation?* Working Paper, January 2002 (downloaded Aug. 15, 2002) <http://gsbwww.uchicago.edu/research/workshops/elo/lerner2.pdf> and Corporate Technology Directory, Editions 1990-2000; [12]*Yankee Group: Video-Game Penetration Grows to 36 Million Households in 2001,* REUTERS NEWS, November 19, 2001 (visited Aug. 30, 2002) <http://about.reuters.com/newsreleases/art_19-11-2001_id785.asp>; [13]Game Makers Hedge Bets in Console Wars, USA Today, November 16, 2001 (visited Aug. 23, 2002) <http://www.usatoday.com/life/cyber/tech/review/games/2001/11/19/game-makers.htm>; [14]DAVID S. EVANS & RICHARD SCHMALENSEE, PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING 170 (1999).

CHAPTER

5 Two-Sided Markets

David S. Evans

ABSTRACT

This chapter addresses the analysis of market definition when the parties involved in an antitrust or merger analysis include one or more two-sided platforms. We discuss how standard market definition measures such as SSNIP tests, diversion ratios, and conditional logit demand analyses have to be modified to account for the unique characteristics of two-sided platforms. We also review how market definition of two- sided platforms was treated in recent US and EC case law.

I. INTRODUCTION

Nightclubs and online dating companies provide services to men and women who want to meet. Credit cards enable consumers who have them and merchants that accept them to transact with one another. Video game platforms, such as Sony PlayStation or Nintendo, provide software tools that enable publishers to develop games and a device on which consumers can play those games. In each case, the business provides a platform that enables two distinct but related groups of customers to obtain value. These "two-sided platforms" or "two-sided markets" are the major form of business organization in many industries.[1]

---

[1] David S. Evans & Richard Schmalensee, *Industrial Organization of Markets with Two-Sided Platforms*, 3 Competition Pol'y Int'l 150, 152 (2007) [hereinafter Evans & Schmalensee, *Industrial Organization*] ("Many diverse industries are populated by businesses that operate 'two-sided platforms.' These businesses serve distinct groups of customers who need each other in some way, and the core business of the two-sided platform is to provide a common (real or virtual) meeting place and to facilitate interactions between members of the two distinct customer groups. Two-sided platforms are common in old-economy industries such as those based on advertising-supported media and new-economy industries such as those based on software platforms and web portals. They play an important role throughout the economy by minimizing transactions costs between entities that can benefit from getting together."). Many platforms have more than two groups of distinct users. *See* David S. Evans & Michael Noel, *The Analysis of Mergers That Involve Multisided Platform Businesses*, 4 J. Competition L. & Econ. 663, 664 (2008) [hereinafter Evans & Noel] ("Many old industries are based on [multisided platforms or] MSPs, ranging from village matchmakers that date from ancient times to advertising supported newspapers introduced in the seventeenth century to payment cards introduced in the mid-twentieth century. However, an increasing number of significant modern businesses are MSPs as a result of technological changes that have drastically lowered the costs and increased the benefits of connecting diverse customer groups on a single platform. These include most Internet-based businesses such as eBay, Facebook, and Google. These businesses are creating new products and services such as social networking platforms and are disrupting existing industries such as advertising-supported media."). This chapter focuses on two-sided cases, but most concepts extend readily to multisided cases.

In antitrust cases involving two-sided platforms, market definition and market power analyses must take into account several economic issues that do not arise in other contexts. The two sides of a platform business are closely linked, with interdependent prices and outputs and intertwined strategies To understand the relevant competitive relationships, one must consider both sides of the platform business.

For example, if an online dating service increased its prices to women, it would attract fewer women to its site. As a result, it would lose male members and, with fewer men, even more women would desert the service. Credit card issuers offer reward programs to persuade more people to use credit cards. As credit card use increases, card issuers can convince more merchants to pay for taking the card.

A platform can earn profits on either side. In practice, two-sided platforms often obtain most of their incremental profits on one side and may provide services to the other side at prices below incremental costs. For example, video game platforms earn most of their profits "on game developers through per-unit royalties on games and fixed fees for development kits and treat the gamers side as a loss leader."[2]

Most standard approaches to market definition, such as the SSNIP (small but significant and nontransitory increase in price) test, diversion ratios, conditional logit demand analyses,[3] and other economic models and formulae, do not apply to two-sided markets without modification, occasionally radical in nature.[4] Many of these "standard tools of antitrust and merger analysis . . . were developed based on the economics of single-sided businesses" and "do not necessarily apply in ways that are material to the analysis of competition that involves multisided businesses."[5] In most cases, that relationship does not hold without significant modification for two-sided platforms. For example, one common approach—using the price-cost margin on one side to assess critical

---

[2] Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. Eur. Econ. Ass'n 990, 990-91 (2003) [hereinafter Rochet & Tirole, *Platform Competition*].

[3] *See*, e.g., United States v. SunGard Data Sys., 172 F. Supp. 2d 172, 186-93 (D.D.C. 2001) (rejecting the government's market definition based on critical loss and SSNIP analyses); FTC v. Tenet Health Care, 186 F.3d 1045, 1050-51, 1053 (8th Cir. 1999) (relying on critical loss analysis).

[4] David S. Evans, *Competition and Regulatory Policy for Multi-Sided Platforms with Applications to the Web Economy*, 2 CONCURRENCES 57 ¶ 28 (2008) (providing an overview of two-sided platforms); Evans & Noel, supra note 1, at 675 (explaining that "the one-sided Critical Loss formulas for conducting the SSNIP test are wrong when applied to two-sided markets"); Evans & Schmalensee, *Industrial Organization, supra* note 1, at 151 ("[P]rofit-maximizing prices may entail below-cost pricing to one set of customers over the long run and, as a matter of fact, many two-sided platforms charge one side prices that are below marginal cost and are in some cases negative. These and other aspects of two-sided platforms affect almost all aspects of antitrust analysis—from market definition, to the analysis of cartels, single-firm conduct, and efficiencies.").

[5] Evans & Noel, *supra* note 1, at 664 ("Each side of the MSP's business influences and constrains its strategies on the other side. Antitrust analysis that focuses on one side of the business in isolation from the other side is incorrect as a matter of economics, and can lead to the wrong answer when indirect network effects are significant and are relevant for assessing the practice at issue.").

loss—tends to understate the effects of a merger on prices compared with the two-sided market formula.[6] Another approach—estimating demand elasticities directly based on a standard one-sided model—tends to overstate the effects of a merger on prices.[7]

This chapter describes how two-sided platforms operate, explains the basic economics of profit-maximization for two-sided businesses, and notes some of the unique aspects of competitive strategies when one of the rivals is two-sided. It summarizes key market definition issues, shows the extent to which the use of standard techniques for two-sided platforms can result in bias, and addresses some basic conceptual issues for markets with two-sided platforms, such as how to measure output and which firms to include in the analysis. Finally, it discusses cases and agency decisions concerning two-sided markets.

II. THE ECONOMICS OF TWO-SIDED PLATFORMS

A two-sided platform provides goods or services to two distinct groups of customers who need each other in some way and who rely on the platform to intermediate transactions between them.[8] Two-sided platforms "minimiz[e] transactions costs between entities that can benefit from getting together[,]"[9] permitting value-creating exchanges to take place that would not occur otherwise.

> The fundamental role of a two-sided platform in the economy is to enable parties to realize gains from trade or other interactions by reducing the transactions costs of finding each other and interacting. Two-sided platforms do this by matchmaking, building audiences, and minimizing costs. Different platforms engage in these activities to different degrees. Software platforms are mainly about minimizing duplication costs, advertising-supported media in mainly about building audiences, and exchanges are mainly about matchmaking. But they all seem to engage in each to some degree. All platforms help reduce costs by providing a virtual or physical meeting place for customers.[10]

---

[6] *Id,* at 682.

[7] *Id.*at 670.

[8] *See* Evans & Schmalensee, *Industrial Organization, supra* note 1, at 151. *See generally* David S. Evans, *The Antitrust Economics of Multi-Sided Platform Markets,* 20 YALE J. ON REG. 325 (2003) [hereinafter Evans, *Antitrust Economics*]; David S. Evans, *Some Empirical Aspects of Multi-Sided Platform Industries,* 2 Rev. Network Econ. 191 (2003); Rochet & Tirole, *Platform Competition, supra* note 2; Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37 Rand J. Econ. 645 (2006) [hereinafter Rochet & Tirole, *Two-Sided Markets*].

[9] Evans & Schmalensee, *Industrial Organization, supra* note 1, at 151; David S. Evans & Richard Schmalensee, Catalyst Code: The Strategies Behind the World's Most Dynamic Companies 53 (Harvard Business School Press, 2007) [hereinafter CATALYST CODE]; see Rochet & Tirole, *Platform Competition, supra* note 2, at 995.

[10] Evans & Schmalensee, *Industrial Organization, supra* note 1, at 158 (explaining that platforms "minimize transaction costs" through "matchmaking, audiencemaking, and . . . elimination of duplication").

For example, newspapers, magazines, and other types of advertiser-supported media use content to attract readers and then sell access to those readers to advertisers. Advertisers need readers; media platforms reduce advertisers' cost of reaching target audiences, facilitating profitable transactions.[11] Online auction websites such as eBay provide platforms enabling sellers of goods or services find and consummate transactions with buyers. Importantly, each customer on one side tends to realize more value when there are more customers on the other side.[12] Shoppers value shopping malls that have more stores that are of potential interest to them while retail stores value shopping malls that have more potential customers.

Two-sided platforms usually perform three core functions to some degree.[13] They serve as matchmakers to facilitate exchange by making it easier for members of each group to find each other (e.g., eBay, MySpace.com, NASDAQ). They build audiences because this makes it more likely that members of a group will find a suitable match (e.g., Google, TiVo, Condé Nast, Fox News). Finally, they provide shared resources and reduce the cost of providing services to both groups of customers (e.g., Windows, Sony PlayStation, Xbox, SAP enterprise software, Linux, Palm OS).

Most two-sided platforms engage in all three core functions. For example, eBay makes it easier for a person who has a particular item and a person who wants a particular item to find each other. It does this in part by attracting a large number of buyers and sellers to its platform thereby making thicker markets for each item. Finally, it provides a common platform for matchmaking that reduces the costs of exchange for all participants.

A key feature of two-sided platforms is the presence of "indirect network effects."[14] Indirect networks effects exist when the value that a customer on one side realizes from the platform increases with the number of customers on the other side. A search platform is more valuable to advertisers if it is more likely that it will reach a larger number of potential buyers. It is more valuable to users looking to buy something if there are more advertisers attracted to the platform because that makes it more likely that the user will see a relevant advertisement.[15] As a practical matter, it is often the strength of these indirect network effects that determines whether being "two-sided" matters enough to have a substantive effect on the results of economic analysis.[16]

---

[11] *See generally* Simon P. Andersen & Stephen Coate, *Market Provision of Broadcasting: A Welfare Analysis*, 72 Rev. Econ. Stud. 947 (2005).

[12] Many two-sided platforms have positive indirect networks for both sides. In the case of advertising-supported media, advertisers value a platform more if it has more viewers but viewers may not value a platform more if it has more advertising; the platform solves this asymmetry in valuation by paying viewers to join the platform in the form of free content. Evans & Schmalensee, *Industrial Organization, supra* note 1, at 155-56.

[13] Catalyst Code, *supra* note 9, at 7

[14] Evans, *supra* note 4, ¶ 8 & n.10 (citing Rochet & Tirole, *Two-Sided Markets*, supra note 8).

[15] See David S. Evans, *The Economics of the Online Advertising Industry,* 7 Rev. Network Econ. 359, 372-75 (2008).

[16] In addition to traditional indirect network effects, externalities in use are also important: in the case of payment cards a product cannot exist unless there is a cardholder and an accepting merchant. *See* Evans, *supra* note 4, ¶ 8 & n.12 (citing Rochet & Tirole, *Platform Competition, supra* note 2).

Two-sided platforms' key feature is the symbiotic relationship between the two sides; the platforms must cater to two different customer groups simultaneously. To establish a two-sided platform, the founders must solve the "chicken-and-egg problem": customers on side A will not participate without customers on side B while customers on side B will not participate without customers on side A. The founders must be able to make credible commitments to one side that if they show up at the platform, the other side will be there as well. In some cases these commitments must be made virtually at the same time. For example, consumers will not carry a payment card unless they know they can use it at merchants' stores, and merchants and the processors who service them will not incur the cost of accepting a card unless they know that enough customers want to use it. In other cases these commitments occur sequentially. A shopping mall must first convince retailers that enough shoppers will come once the retailers locate there. Then the mall owner tries to attract shoppers to its completed mall.[17]

A related feature of two-sided platforms is the need to "balance" the demands of the two sides. In setting prices, a two-sided platform needs to consider that charging a higher price to side A will result in fewer A's using the platform which in turn will result in fewer B's getting value from the platform. It also needs to consider that there must be enough A's to be of interest to the B's and enough B's to be of interest to the A's in most cases. Setting store rental prices too high at malls could result in shoppers not having access to the stores they most care to go to.

A two-sided platform's profit-maximizing calculus is thus more complex than that of a traditional business. The two-sided platform needs to consider the demands of both sides, the interrelationships between these demands, the costs directly attributable to each side, and the costs of running the platform. Many platforms consider both access fees to join the platform and usage fees to interact on the platform in establishing prices that attract the optimal volume of transactions on each side of the platform.

One side of a two-sided platform usually gets a better deal or receives service for free. For example, searchers do not pay search engines, advertisers do. Buyers do not pay a fee on many online transaction platforms, sellers do. Merchants, rather than cardholders, pay for most of the cost of payment card schemes. Retail stores, not shoppers, pay to get access to a shopping mall. End users pay for most software platforms, but software developers often get access to code for free. Women often get discounts from dating services, while men seldom do.

There are many practical reasons for price differences between the two sides of a platform. A two-sided platform has to attract enough members on each side to have sufficiently thick markets. Most business-to-business exchanges failed in the early 2000s because, even though many buyers

[17] See David S. Evans, *How Catalysts Ignite: The Economics of Platform-Based Start-Ups, in* Platforms, Markets And Innovation (A. Gawker ed. forthcoming 2009) (discussing solving the start-up problem for two-sided platforms); *see also* CATALYST CODE, supra note 9, at 89-91 (discussing two-sided platforms' pricing and commitment); Andrei Hagiu, *Pricing and Commitment by Two-Sided Platforms*, 37 (3) Rand J. Econ. 720 (2006) (same).

had signed on, there were not enough sellers to make the exchange attractive.[18] A two-sided platform may therefore find that it has to keep prices very low to one side to attract enough participants to make the platform valuable to the other side. The side that is needed more or that is harder to get receives a price break; conversely the side that gets the most value out of access to members of the other side bears more of the cost. Economists have shown that the profit-maximizing prices for two-sided platforms can result in one side getting a price that is less than the incremental cost of providing a customer on that side and sometimes less than zero.[19]

Economic theory predicts that there will typically be a close relationship between prices and incremental costs for traditional single-sided businesses. Competition drives prices down to marginal cost[20] and prices and marginal costs tend to move in the same direction. Deviations between price and marginal cost are often taken as an indicator that firms have market power.[21] In two-sided businesses, however, there is not a close relationship between prices and marginal costs for the goods or services provided on a particular side. Because the conditions for profit maximization depend on demand elasticities and marginal costs for both sides and the cross-side network effects,[22] in theory, a profit-maximizing two-sided platform will respond to an increase in cost of side 1 with an increase in price on side 2 and no increase in price on side 1.

Some two-sided platforms set prices below marginal cost on one side to attract customers to whom they charge prices above marginal cost on the other side. Competition between platforms drives overall profits down. However, it is common to have a competitive equilibrium in which all of the firms have adopted the same skewed pricing structure that recovers most of the costs from one side while favoring the other side. The major payment card schemes in the United States, for

---

[18] See Bob Tedeschi, *Investing; How Killer B-to-B's Went Into A Tailspin*, N.Y. Times, May 7, 2000; Hazel Ward, *B2B Exchanges Fail to Deliver*, ComputerWeekly.com, Nov. 29, 2001, http://www.computerweekly.com/Articles/2001/11/29/183994/b2b-exchanges-fail-to-deliver.htm; Kate Hanaghan, B2B Exchanges—What *Happens Now?,* Silicon.com, Jan. 23, 2002, http://software.silicon.com/webservices/0,39024657,11030680,00.htm.

[19] *See, e.g.,* Rochet & Tirole, *Two-Sided Markets, supra* note 8, at 659 ("[I]t is quite common for a platform to charge below-cost (perhaps zero) prices to one side and high prices to the other."); Mark Armstrong, *Competition in Two-Sided Markets*, 37 Rand J. Econ. 668, 682 (2006) (noting that in some cases prices to one side of the two-sided platform are "below cost").

[20] See Evans, *supra* note 4, ¶ 14 (citing Paul A. Samuelson & William D. Nordhaus, Economics (18th ed. McGraw-Hill/Irwin 2005)).

[21] *See* U.S. Dep't of Justice & Fed. Trade Comm'n, Horizontal Merger Guidelines, § 0.1 (1992) [hereinafter MERGER GUIDELINES], reprinted in 4 Trade Reg. Rep. (CCH) ¶ 13,104 ("Market power to a seller is the ability profitably to maintain prices above competitive levels for a significant period of time."); *see also* Dennis W. Carlton, *Market Definition: Use and Abuse*, 3 (1) Competition Pol'y Int'l 2, 5 (2007) ("Roughly speaking, 'market power' means that the industry's behavior deviates from perfect competition. One standard definition of market power is the ability to set price profitably above the competitive level, which is usually taken to mean marginal cost.").

[22] *See* Rochet & Tirole, *Platform Competition, supra* note 2, at 996-97.

example, all have pricing structures in which merchants pay for the much of the cost of the system while cardholders are subsidized through rewards and the float on their balances.[23]

A common condition for economic efficiency is that "the competitive price equals marginal cost of production."[24] Nonetheless, this condition does not necessarily hold in the case of two-sided platforms in part because there is joint production of services to the two sides.[25] For example, a shopping mall provides services jointly to retailers and shoppers. In some cases, the product cannot exist without both sides' participation: A credit card provides transaction services only if there is a person who uses the card for payment and a merchant who takes the card for payment.[26]

These considerations have implications for the analysis of market power. The analysis of either side of a two-sided platform in isolation yields a distorted picture of the business. The low-margin side will tend to appear highly competitive in the sense that prices are not significantly higher than marginal cost. The high-margin side will tend to appear much less competitive because prices are much higher than marginal cost. Competition between platforms may result in squeezing the margins down on the high-margin side but will not fundamentally alter the skewed pricing structure which results from all firms' seeking to perform the same balancing act. To analyze market power and to understand competitive and anticompetitive strategies involving two-sided platforms one has to consider the interplay between the two sides.

III. MARKET DEFINITION IN CASES INVOLVING TWO-SIDED PLATFORMS

The purpose of market definition at its broadest level is to provide context for an antitrust inquiry by limiting the analysis to a manageable group of relevant companies and products. The relevant companies and products are the ones that could potentially limit the effects of the competitive practice at issue. The availability of demand-side substitutes limits the ability of a firm to raise prices and therefore harm consumers.[27] A firm's rivals may react to the practice at issue by countering

---

23 Evans, supra note 4, ¶ 15 & n.18 (citing David S. Evans & Richard Schmalensee, *The Economics of Interchange Fees and Their Regulation*: An *Overview*, in Interchange Fees in Credit and Debit Card Industries: What Role For Public Authorities? (Santa Fe: Federal Reserve Bank of Kansas City, May 2005)).

24. *See* Dennis W. Carlton & Jeffrey M. Perloff, Modern Industrial Organization 70 (4th ed. 2005).

25 This point was first observed for the case of payment cards in a seminal paper that is a precursor to the two-sided economics literature." Evans, *supra* note 4, ¶ 16 n.19 (citing William F. Baxter, *Bank Interchange of Transactional Paper: Legal and Economic Perspectives*, 26 J. L. & ECON. 541, 544 (1983)).

26 *See* Baxter, *supra* note 25, at 544 ("The mechanics of transactional services require that for every transaction in which a purchaser becomes a maker of a check, there must be one- and precisely one-transaction in which a merchant becomes a payee; similarly, each use of a credit card by a card holder must be matched by precisely one act of acceptance of the card (or, more accurately, the paper that the card generates) by a merchant."). *See generally* Nat'l Bancard Corp. v. Visa U.S.A. Inc., 779 F.2d 592, 594 (11th Cir. 1986) (discussing "the workings of the bank credit card industry").

27 *See* Merger Guidelines, *supra* note 21, § 1.

its harmful effects or reinforcing them.[28] Market definition provides a screen that limits detailed antitrust scrutiny to those situations in which one cannot count on competition to counter the effect of a practice on price, output, innovation, or other measures of competitive performance.[29]

The business practice at issue must be considered in the context of both sides of the platform. A major competitive constraint on practices that affect one side of a two-sided platform comes from the impact of these practices on the platform's other side. A business practice that increases price and thus reduces the number of customers on one side of a two-sided platform will result in a decrease in demand on the other side for the reasons discussed above. Therefore, in assessing whether a business practice could lead to a price increase on one side of the platform, one would have to take into account both the reduction in demand on that side of the platform and the subsequent negative feedback effects on the other side.[30]

The newspaper industry illustrates this point. Newspapers provide content that attracts readers and sell access to these readers to advertisers. Most newspapers earn the preponderance of their profits from advertisers.[31] Most major daily newspapers are sold at a price that roughly covers the incremental cost of printing and distributing those papers but does not contribute to the cost of providing the content.[32] When newspapers establish their profit-maximizing prices and production levels, they necessarily consider the interrelated impacts on the reader and advertiser sides.[33] If a newspaper were to increase the cover charge to its readers, the number of readers would decrease, resulting in the loss of advertising revenue.[34] Any analysis of business practices would hazard missing important effects if it did not consider the interrelated sides of the business.[35]

Courts and agencies have taken differing approaches to defining relevant markets in newspaper industry cases: some have considered the feedback effects between readers and advertisers, while others have focused on just one side of the platform. For example, in *United States v.*

[28] *Id,* § 2.

[29] *Cf.* Carlton, *supra* note 21, at 22-23 (listing common mistakes in market definition).

[30] Traditional market definition analysis focuses on the role of substitutes in constraining price—it does not focus, as it must in two-sided platform markets, on the role of complementary products (the two interlinked platform products).

[31] Evans & Schmalensee, *Industrial Organization, supra* note 1, at 156.

[32] *Id.; see* Roger D. Blair & Richard E. Romano, *Pricing Decisions of the Newspaper Monopolist,* 59 S. Econ. J. 721, 721 & n.2 (1993) (confirming that newspaper cover price is below marginal cost and the sales revenue does not cover the cost of ink and paper used in producing the newspaper) (citing studies showing that, in the 1980s, "a daily newspaper with a circulation of 250,000 typically generated circulation revenue equal to only 75-80 percent of the cost of ink and newsprint").

[33] *See* Blair & Romano, *supra* note 32, at 722 ("Given the feedback effect of circulation on advertising demand, the publisher must be particularly concerned about the circulation price, which is not always under the publisher's control absent vertical integration. Given the feedback effect of circulation on advertising demand, the publisher must be particularly concerned about the circulation price, which is not always under the publisher's control absent vertical integration.").

[34] *See id.*

[35] See Chapter 5 for a discussion of the newspaper industry.

Times-Picayune Publishing,[36] the Supreme Court recognized that "every newspaper is a dual trader in separate though interdependent markets; it sells the paper's news and advertising content to its readers; in effect that readership is in turn sold to the buyers of advertising space."[37] Taking into account both sides of the platform business—its readers and its advertisers—the Court concluded that Time-Picayune Publishing did not occupy " a 'dominant' position in the newspaper advertising market in New Orleans."[38] By contrast, the U.K. competition authorities have acknowledged "the two-sided nature" of the publishing industry in some cases, while considering "only one side of the market" in other cases.[39]

A. Defining Relevant Products

Two-sided platforms serve two distinct sets of customers who receive different but related products or services. In some cases, the products or services the two sides receive are sufficiently similar that they can be combined into a simple metric that can be compared across firms and manipulated to calculate market shares and the Herfindahl-Hirschman Index (HHI). For example, online matchmaking businesses for men and women charge men and women similar prices and provide similar services to both. In a merger of two online matchmaking businesses, each firm's total revenues would be an accurate measure of their market positions.[40] However, market definition would also need to take into account the special competitive constraints on the conduct and pricing of two-sided platforms. If the combined firm were to raise prices to both sides, demand for its services would decline on each side. As a result, each side would value the matchmaking venue less because fewer potential matches would be available.[41]

Most two-sided platforms, though, provide different products or services to the two sides. For example, software platforms such as Sony PlayStation provide game developers with software code to help them write games and supply users with game consoles and software enabling them to play games. Although game users and game developers are relying on the same code and hardware, they are paying different prices and are receiving different services. No single market share metric accurately summarizes the position of Sony or of competing video console makers. To understand market dynamics, one must consider both the competitors' shares of video console sales and their shares of game sales.

---

[36] 345 U.S. 594 (1953).

[37] *Id.* at 610.

[38] *Id.* at 610-11, 627-28 (overruling the district court's finding of dominance).

[39] *See* John Wotton, *Are Media Markets Analyzed as Two-Sided Markets?,* 3 Competition Pol'y Int'l 237, 240-44 (Spring 2007) (discussing cases).

[40] *See* Evans, *Antitrust Economics, supra* note 8, at 360-361. The two-sided considerations for implementing the SSNIP test are still relevant in this case because a symmetric price increase will not only lower the direct demand for each side but also lower the indirect demand through the feedback effects. Evans & Noel, *supra* note 1, at 666.

[41] Evans & Noel, *supra* note 1, at 667.

B. Considering Competitive Constraints on Both Sides

The competitive constraints on both sides of the platform should be taken into account. Since any business practice a two-sided platform engages in could result in an attempt to raise prices on either side, the ability to do so on either side affects the likelihood that the business practice will result in higher prices and profits. Consider alleged predatory pricing by a platform that is in competition with other platforms that all serve the same sides. As many contributors to the two-sided economics literature have observed, it is not meaningful to talk about predatory pricing on one side since it may in fact be profitable—and socially efficient—for all firms to charge less than marginal cost on that side.

Analyzing pricing on just one side could therefore lead to a false positive or to a false negative. A platform might reduce its price drastically on one side to drive a rival out of business but then choose to recoup its losses by raising price on the other side. One would need to examine whether prices have been lowered enough to make it unprofitable for competing platforms to operate at the margin (i.e., whether it would be profitable to carry out a joint incremental expansion of both sides of the platform). Under some legal tests one would also need to examine whether it is plausible that the preying platform could recoup its losses. A sound analysis would have to consider whether it is possible for the platform to recoup as a whole (i.e., across both sides) and should not focus on any particular side.

Consider the payment card industry. A card scheme could enter into a massive promotion campaign to persuade merchants to take its card exclusively. An analysis of the merchant side only might reveal that the card scheme has not recouped its losses. But the card scheme might also raise fees to cardholders. Taking both sides into account, it does recoup its predatory losses.[42]

C. Accounting for Diverse Business Models

Two-sided platforms coexist and compete with other business models to fulfill customer needs. Market definition must consider the diverse ways in which a two-sided platform may face rivalry, taking into account the market participants' reactions to price changes. These reactions are more difficult to predict when the firms are following different business models.

First, a two-sided platform may face single-sided competition on one or both sides. For example, a shopping mall developer faces competition from single stores for the attention of shoppers, as well as from real-estate investors that rent single-store locations. The degree to which these single-sided alternatives constrain the two-sided mall's conduct is an empirical question.

Second, a two-sided platform may compete with a three-sided platform. The three-sided rival produces another product which could, for example, have below-cost prices on both sides served

[42] *See* Julian Wright, *One-Sided Logic in Two-Sided Markets,* 3 (1) Rev. Network Econ. 44, 48 (2004) (explaining why a two-sided platform's below-cost pricing does not indicate predation).

by the two-sided platform. A two-sided platform is particularly vulnerable to competition by a three-sided platform that uses its third side to subsidize both of the other sides. This asymmetric competition is potentially lethal to a platform that does not provide the third side. A multi-sided platform can therefore use "envelopment" to challenge platforms that provide a subset of the services it provides.[43] It adds another group of customers to the platform and uses revenues from this group of customers to lower the price—possibly to zero—of the key profit-generating side of the other platform.

For example, Google gives away office productivity software that competes with Microsoft's software to users and software developers.[44] By doing so, Google attracts customers and profits by selling access to those customers to advertisers.[45] Google can use its advertising revenue to compete with Microsoft in software. Competition from this advertising-supported software model has led Microsoft to enter into advertising to ensure that it also has a stream of advertising revenue available.[46]

Third, a two-sided platform may compete with a business that has vertically integrated into one of the sides and is therefore effectively single-sided.[47] Many two-sided platforms provide services to end users and to producers of complementary products that their end users value. Some businesses may decide to produce the complementary products themselves. Video game console makers used to produce their own games and it was not possible to use third party games with the consoles.[48] Apple has chosen to make its iPod work only with its own media store iTunes while Sony has chosen to make its line of mp3 and media players work with multiple sources of media.

---

[43] *See* Evans, *supra* note 4, ¶ 25 & n.23 (citing Thomas R. Eisenmann, Geoffrey Parker & Marshall W. Van Alstyne, *Platform Envelopment* 1 (Working Paper 2007), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=996852).

[44] *See* Evans, *supra* note 4, ¶ 26.

[45] *See* Fred Vogelstein, *Gates vs. Google: Search and Destroy*, Fortune, Apr. 18, 2005 (Google "does not sell its search products [but] makes its money from the ads that accompany its search results"); *see also* Charles H. Ferguson, *What's Next for Google*, 108 (1) Tech Rev. 38 (2005) (discussing competition between Google and Microsoft).

[46] Microsoft's investment in advertising has included the launch of Microsoft adCenter, the company's online advertising platform, and the purchase of aQuantive, a family of digital marketing service and technology companies, for $6 billion. Press Release, Microsoft, Seventh Annual MSN Strategic Account Summit Celebrates Advertisers' Role in Microsoft Media Network Vision (May 3, 2006), http://www.microsoft.com/presspass/press/2006/may06/05-03SAS7PR.mspx; Press Release, Microsoft, Microsoft Acquires aQuantive (May 18, 2007), http://advertising.microsoft.com/uk/microsoft-aquantive-announcement.

[47] The history of software platforms shows that firms can start out vertically integrated by providing both the software platform and its key complement such as "hardware, applications, and peripherals." See David S. Evans, Andrei Hagiu & Richard Schmalensee, Invisible Engines: How Software Platforms Drive Innovation and Transform Industries 301 (2006) [hereinafter Invisible Engines] (explaining that the PalmPilot began as an integrated bundle).

[48] *See* Invisible Engines, *supra* note 47, at 301 ("[I]n video games both Sony and Microsoft acquired several high-profile game developers before releasing their consoles so that both end users and other game developers could reliably expect their consoles to have a number of high-quality games.").

Finally, a two-sided platform may compete with a series of businesses that, when combined, provide similar services to the same two groups of customers. For example, American Express and Visa compete for cardholders and merchants. American Express is an integrated two-sided platform that performs all of the necessary steps for meeting cardholder and merchant demand. By contrast, Visa is a collection of separate entities that—when combined—perform services similar to those provided by American Express. Working from the merchant level to the cardholder level, there are acquirers that contract to take card receivables, merchant processors that handle the actual data processing and switching services, Visa that stands at the center of this platform as the switch for transactions, card processors that handle data processing for banks, and banks that issue cards to consumers.

Each of these market constellations results in different forms of competitive constraints and possible counter strategies from rivals. The market boundaries that separate relevant from irrelevant constraints are particularly complex. That fact cautions against placing great weight on mechanical approaches to market definition that cannot account for competitive nuances and complexities.

A natural way to map out the contours of the market is to identify the groups of customers served by the subject of the inquiry and its likely rivals, and then identify the various businesses that serve these customers. The digital media business provides an example. Suppose one needs to define the relevant product market for a media player such as Adobe's Flash. As of 2007, Adobe has three groups of customers— end users who receive its player for free; enterprises that pay for its server; and content providers that use its tools to create content.[49] These three groups of customers are served to various degrees by (1) Apple's iPod/iTunes, a vertically integrated platform with its own media player; (2) Microsoft's Windows Media Player technologies included both in its client and server software; (3) RealPlayer, a stand-alone product for clients and servers and the base for a media store that competes with iTunes; (4) software vendor Autodesk that provides tools to content providers to make their content available in various formats including Windows Media, QuickTime, or MPEG-2; and (5) various other businesses including Sorenson, Canopus, and others.[50] Not all these entities will necessarily be part of a properly defined relevant market. However, to define relevant markets in cases involving two-sided platforms, one must consider multiple types of potential competitors that may constrain the pricing and other strategies of the subject of the inquiry.

IV. TWO-SIDED CRITICAL LOSS ANALYSIS AND SSNIP TESTS[51]

The standard approaches to market definition do not apply to two-sided markets without significant modification. Even the two-sided versions of these approaches should be used with caution.

---

[49] *Id,* chapter 7.

[50] *Id,* chapter 8.

[51] *See* Chapter 1 for a discussion of critical loss analysis and SSNIP tests in one-sided markets.

The price-cost margin, also known as the Lerner Index of market power, is often used in merger analysis.[52] The Lerner Index equals "the difference between price and marginal cost" expressed as a percentage of price.[53] When a firm has maximized its profits, the Lerner Index equals the inverse of the elasticity of demand.[54] The elasticity of demand "is defined as the percentage change in quantity that results from a 1 percent change in price."[55] For example, a 50-percent profit margin corresponds to an elasticity of demand of two.[56]

> If the elasticity of demand is very high (a large negative number), then the [demand] curve is said to be very *elastic*. With a very elastic demand, a small price change induces a very large change in the quantity demanded. If the elasticity is low (a number between –1 and 0), the demand curve is *inelastic*, and a price change of 1 percent has relatively little effect on the quantity demanded.[57]

The Lerner Index—calculated in this way for a single product—must be considerably modified for use with two-sided platforms.[58] This section explains why the Lerner Index and other traditional merger analysis techniques must be modified and describes the two-sided version of these tests.

---

[52] *See* Carlton & Perloff, *supra* note 24, at 93.

[53] *Id.*

[54] *See id.* at 92-93 ("[T]he price-cost margin depends only on the elasticity of demand the monopoly faces."). To maximize profits, a firm expands output to the point at which "the extra revenue from selling one more unit just equals the extra cost of producing that last unit of output. That is, profit is maximized where *marginal revenue equals marginal cost.*" Id. at 91 (emphasis in original). When incremental revenue is greater than incremental cost, "total revenue increases as output expands," but when incremental revenue is less than incremental cost, "total revenue falls as output expands." *Id.* at 90. When profits are maximized, the firm's percentage profit margin is inversely proportional to the elasticity of demand for the firm's product. *See id.* at 92-93 ("The higher the elasticity of demand, the closer is the monopoly price to the competitive price. Therefore, the key element in an investigation of market power is the price elasticity of demand. Where the elasticity of demand is relatively inelastic, a monopoly markup may be substantial[.]"). The Lerner Index is based on various assumptions that may not hold true in practice and, as shown below, do not in fact hold true in the case of multisided platforms.

[55] *Id.* at 92. When marginal and average variable costs are equal, the Lerner Index equals the contribution margin—gross margin (revenue minus variable cost) expressed as a percentage of revenue—which is often available from reported accounting data. *See* Evans & Noel, *supra* note 1, at 686.

[56] If the profit margin is 50% or ½, its inverse is one divided by ½, which equals 2.

[57] See Carlton & Perloff, *supra* note 24, at 92 (emphases in original).

[58] *See* Evans & Noel, *supra* note 1, at 676 ("The one-sided Lerner Index is incorrect for MSPs [multisided platforms] and relying on it significantly overestimates the true short-run own-price elasticity of demand and over-states Actual Loss."). There is no necessary theoretical relationship between marginal revenue and marginal cost on each side in isolation for the same reason there is no necessary relationship between price and marginal cost. *See* Evans & Noel, *supra* note 1, at 666. For many two-sided platforms, marginal revenue exceeds marginal cost on one side, but is less than marginal cost on the other side. It is possible to extend the single-sided Lerner Index to an analogous concept for two-sided platforms. *See* Evans & Noel, *supra* note 1, at 663 *passim* (exploring approaches).

A. The SSNIP Test[59]

The SSNIP test underlies market definition in merger analysis.[60] The antitrust agencies begin with each product (narrowly defined) produced or sold by each merging firm and ask what would happen if a hypothetical monopolist of that product imposed at least a "small but significant and nontransitory" increase in price [SSNIP], but the terms of sale of all other products remained constant. If, in response to the price increase, the reduction in sales of the product would be large enough that a hypothetical monopolist would not find it profitable to impose such an increase in price, then the [agencies] will add to the product group the product that is the next-best substitute for the merging firm's product.[61]

The agencies then ask

> [t]he price increase question . . . for a hypothetical monopolist controlling the expanded product group. In performing successive iterations of the price increase test, the hypothetical monopolist will be assumed to pursue maximum profits in deciding whether to raise the prices of any or all of the additional products under its control. This process will continue until a group of products is identified such that a hypothetical monopolist over that group of products would profitably impose at least a "small but significant and nontransitory" increase, including the price of a product of one of the merging firms. The [agencies] generally will consider the relevant product market to be the smallest group of products that satisfies this test.[62]

In most cases, the agencies "will use a price increase of five percent lasting for the foreseeable future," but may "use a price increase that is larger or smaller than five percent" depending on the industry involved.[63] In one-sided context, the agencies apply the SSNIP test through calculations based on the Lerner Index or through direct statistical estimation of demand relationships.[64]

A two-sided platform "would like to find the prices for each side that maximize its profits."[65] As it adds customers on one side through, for example, an acquisition of a competing firm, it would ordinarily adjust prices to both sides. The standard one-sided SSNIP test would ignore the prices for the two-sided platform's other product, as well as the feedback effects between the two sides.[66]

---

[59] This section compares two-sided and one-sided SSNIP tests. The one-sided SSNIP test's recognized deficiencies are also true in the two-sided context. See Chapter 1 for a full discussion of one-sided SSNIP test.
[60] *See* Merger Guidelines, *supra* note 21, § 1.11.
[61] *Id.*
[62] *Id.*
[63] *Id.*
[64] See Chapter 1 of this volume for a discussion.
[65] Evans & Noel, *supra* note 1, at 665.
[66] *See id.* at 667.

The application of a one-sided SSNIP test in cases involving two-sided platforms can result in a "relevant market" that includes just one side of a two-sided platform, even though that other side imposes significant constraints that would temper a price increase.[67] It would place key aspects of profit maximization and strategic interactions outside the "relevant market."[68] "[T]he market would be drawn too narrowly and estimates of market concentration too high, because the standard approach fails to consider the tempering effects on price coming from the other side."[69]

To illustrate this effect, consider a merger between two heterosexual dating clubs. The standard one-sided SSNIP test would focus on one side of the platform, ignoring the other side and resulting in an overly narrow market definition.[70] Using the standard one-sided SSNIP test, the change in profits of the hypothetical monopolist in response to an increase of the admission fee for men has two components: the monopolist gains from higher admission prices, but loses from the lower volume of sales of dating services to men. However, the feedback effects between the two sides amplify the monopolist's losses from the decrease in the demand for dating services from men.[71] As fewer men access the club, fewer women would be interested in joining the club.[72] The decrease in the number of women would in turn reduce the number of men, which would lead to a further decrease in the number of women, and so on.[73] A price increase deemed profitable under the one-sided SSNIP test may turn out to be unprofitable under the two-sided SSNIP test.[74]

A two-sided extension of the SSNIP test could address these issues at least in theory. The analyst would start with the two-sided platform at issue and add its closest substitutes.[75] The analyst would then ask whether the combined platforms could increase the weighted average price by a given percentage and thereby reduce the weighted average output of the platforms.[76] At each step, the hypothetical monopolist would be expected to "reoptimize prices across sides and across platforms," considering the feedback effects.[77] Additional platforms can be added. This approach becomes more difficult to implement when a two-sided platform faces diverse competitors with different business models.[78]

---

67 *Id.*
68 *See id.*
69 *Id.*
70 *See id.*
71 *See* Evans & Schmalensee, *Industrial Organization, supra* note 1, at 159.
72 *See id.*
73 *See id.*
74 *See* Evans & Noel, *supra* note 1, at 669-70 ("Because the analyst's estimate does not account for feedback effects, the full impact of the price increase on demand is underestimated. As a result, antitrust markets necessarily will be defined too narrowly, and merger analysis will overstate the increase in market power of merging parties and overstate the predicted unilateral price effects of the transaction.").
75 Evans & Noel, *supra* note 1, at 669-78.
76 Evans & Noel, *supra* note 1, at 669-78.
77 *See* Evans & Noel, *supra* note 1, at 674.
78 Evans & Noel, *supra* note 1, at 669-78.

There are "many reasons to be weary of mechanical market definition exercises such as the SSNIP test," which "do not go away"—and in some cases are exacerbated—when two-sided platforms are involved.[79] For example, should one include both sides of the [two-sided platform] business . . . in the market definition of just one side? The answer can be difficult and depends on the situation at hand."[80] In discussing the SSNIP test, the *Merger Guidelines* refer to "adding or removing 'substitute products.'"[81]

> The difficulty is that the sides of a given [two-sided platform] represent highly complementary products, and pricing decisions on both sides critically affect the [platform's] profitability. Yet the SSNIP test is silent on handling complementary products. We believe if the two sides are very highly complementary and closely linked—for example, if the MSPs facilitates transactions between the groups that occur in fixed proportions—and MSPs in an industry all tend to serve the same two sides, then it can be reasonable to include both sides in the market definition and the "transaction" as the product. However, in other industries MSPs may all cater to the same side *A* customers but cater to very different kinds of side *B* customers. If the antitrust concern centered around the side *A* business, then including both sides of all MSPs that share the *A* side in the market definition would open a Pandora's box of unrelated "*B*" types that make no sense under a single coherent market definition. Then the market may need to be defined on the basis of side *A* only, but with the critical understanding that the *B* sides are an important constraint on behavior and that the formulas presented here must be used to account for its influence.[82]

Another issue with the SSNIP test is that "we would expect a hypothetical monopolist of several platforms to reoptimize prices across sides and across platforms just as we would expect a hypothetical monopolist to reoptimize across products in a one-sided world."[83] The *Merger Guidelines* suggest that the price increase deemed significant for market definition purposes will generally be uniform across all products within the definition (for example, 5 percent), and "the terms of sale of all other products [outside the proposed antitrust market] are held constant" (that is, 0 percent price increase). In the case of [two-sided platforms], one could instead imagine price increases that differ across sides or platforms, or consider a kind of 5 percent quantity-weighted "average" price increase across sides, allowing the hypothetical platform monopolist to reoptimize relative prices.[84]

[79] *See* Evans & Noel, *supra* note 1, at 674.
[80] *Id.*
[81] *Id.*
[82] *Id.*
[83] *Id.*
[84] *Id.*

B. Critical Loss Analysis Based on the Lerner Index

Critical loss analysis is a user-friendly technique for implementing the SSNIP test.[85]

> It compares "Critical Loss" (*CL*)—the percentage loss in quantity of a hypothetical monopolist's products that would be exactly enough to make an *X percent* price increase in the price of all its products unprofitable—to "Actual Loss" (*AL*)—the predicted percentage loss in quantity that the monopolist would suffer if it did increase prices on all its products by *X* percent. . . . A relevant market is found when Actual Loss equals Critical Loss for a hypothetical monopolist of the given set of products in the proposed antitrust market. If Actual Loss exceeds Critical Loss, the relevant market is expanded to include more substitutes. Otherwise, it is contracted.[86]

Critical loss is based on the incremental profit margin, while actual loss is based on the responsiveness of demand to changes in price of the product itself as well as that of substitutes.[87] This method "has won significant appeal both because of its simplicity and because of its easy measurement of inputs" based on available data.[88]

A common back-of-the-envelope calculation for merger inquiries uses the Lerner Index to estimate the elasticity of demand for actual loss.[89] However, the traditional Lerner Index for one side of a two-sided platform does not consider the role of the other side of the platform in constraining price.[90] Suppose the analyst observes a profit margin of 25 percent. For a one-sided business, this profit margin would imply that the elasticity of demand facing the firm was four.[91] One could infer

---

[85] *See* Evans & Noel, *supra* note 1, at 668; see also Barry Harris & Cento G. Veljanovski, *Critical Loss Analysis: Its Growing Use in Competition Law,* 5 E.C.L.R. 213 *passim* (2003) (discussing the use of critical loss analysis in antitrust law); James Langenfeld & Wenqing Li, *Critical Loss Analysis in Evaluating Mergers,* 46 (2) ANTITRUST BULL. 300 (Summer 2001) ("The antitrust agencies and economists have increasingly focused on concepts of "critical elasticity of demand" or "critical loss analysis" to define antitrust markets and evaluate the competitive effects of proposed mergers."). *See generally* Barry Harris & Joseph J. Simons, *Focusing Market Definition: How Much Substitution is Necessary?,* 12 RES. L. & ECON. 207 (1989) (discussing the use of the critical loss concept "in practice"); Michael L. Katz & Carl Shapiro, *Critical Loss: Let's Tell the Whole Story,* 17 ANTITRUST 49 (Spring 2003) (discussing market definition using critical loss analysis); Daniel P. O'Brien & Abraham L. Wickelgren, A *Critical Analysis of Critical Loss Analysis*, 71 ANTITRUST L. J. 161 (2003) (explaining economics of critical loss).

[86] Evans & Noel, *supra* note 1, at 668.

[87] *See id.*

[88] *Id.* at 669.

[89] See Chapter 1 of this book.

[90] Evans & Noel, *supra* note 1, at 666; see id. at 676 ("The one-sided Lerner Index is incorrect for [two-sided platforms] and relying on it significantly overestimates the true short-run own price-elasticity of demand and overstates Actual Loss.").

[91] If the elasticity of demand is the inverse of the profit margin, and the profit margin is 25%, then the elasticity of demand is the inverse of 25%, 1/.25, or 4.

from this relatively high elasticity of demand that significant substitutes constrain the ability of the firm to increase prices.

For a two-sided business, a 25-percent profit margin would be consistent with less elastic demand, because the two-sided business would also have to consider the effect of raising prices on revenue from the other side and the resulting feedback effects. Thus a demand elasticity of less than four plus positive feedback effects would result in the 25-percent profit margin. The use of the Lerner Index will tend to overstate the elasticity of demand and understate the extent to which a firm—or the putative hypothetical monopolist—could raise prices.[92] All else equal, this mistake results in underestimating the market power on one side of the two-sided platform and defining markets too broadly on that side.[93]

It is possible to extend the simple version of actual loss analysis to two-sided platforms. The analyst would start with the two-sided platform at issue and consider substitutes.[94] The analyst would have to estimate the feedbacks between the two sides in addition to the profit margins.[95] Table 1 provides examples of the impact of feedback effects on actual loss estimation in cases involving many competing two-sided platforms.[96] It shows that the decision to expand or shrink the market definition depends on the assumed feedback effects between side A and side B (i.e., side A's effect on side B and side B's effect on side A). A decrease in the number of customers on side A as a result of the price increase on side A causes a decrease in the number of customers on side B, which in turn reduces the demand on side A.[97]

C. SSNIP Tests Based on Demand Estimation

With enough data, time, and resources, one can implement a more rigorous version of the SSNIP test by estimating the relevant demand elasticities.[98] A typical one-sided analysis would consider potential substitutes for the products on that side, but would not account for the feedback effects between the two sides.[99] The traditional econometric models' estimates of the elasticity of demand

---

[92] Evans & Noel, *supra* note 1, at 676.

[93] Evans & Noel, *supra* note 1, at 670, 689. One should not, however, lose sight of the most fundamental mistake: leaving out the second side of a two-sided market altogether.

[94] Evans & Noel, *supra* note 1, at 669-78.

[95] Evans & Noel, *supra* note 1, at 669-78. In performing a preliminary calculation based on readily observed data, the analyst might have little basis to estimate the magnitude of the feedback effects.

[96] To simplify the computation of the Critical Loss of output on side A, we assume that the two sides of the hypothetical monopoly platform are tied in fixed proportions. An example of such a platform is a transaction market like credit card services where a transaction takes place between a customer on side A (card holder) and a customer on side B (merchant). With the assumed level of profit margins of 25% on both sides, the critical loss in output on side A is 17%.

[97] *See* Evans & Schmalensee, *Industrial Organization, supra* note 1, at 159.

[98] See Chapter 1 of this volume.

[99] See Chapter 1 of this volume.

TABLE 1 Impact of Feedback Effects on Actual Loss Estimation in Two-Sided Platforms, Assuming a Price Increase of 10% on Side A

Profit Margin on side A	Elasticity of demand on side A	Feedback from side A to side B	Feedback from side B to side A	Actual Loss in Output on side A	Critical Loss In Output on side A	Expand / Shrink
25%	4	2	1	-40%	17%	Expand
25%	4	2	1.1	-33%	17%	Expand
25%	4	2	1.25	-27%	17%	Expand
25%	4	2	1.3	-25%	17%	Expand
25%	4	3	1.11	-17%	17%	Market Defined
25%	4	3	1.25	-15%	17%	Shrink
25%	4	3	1.3	-14%	17%	Shrink
25%	4	3	1.3	-14%	17%	Shrink

for the product in question would be too small, because the models would not account for the feedback effects from the other side of the platforms.[100] These estimates will tend to yield an overly narrow one-sided relevant market.[101]

It is possible, however, to conduct a statistical analysis of demand that recognizes the two-sided nature of the market. Several empirical studies have done so.[102] For example, Rysman studies the feedback loop in the market for yellow pages.[103] He estimates "two demand curves simultaneously": . . . a consumer demand for directory usage as a function of [the demand for] advertising" and "an inverse advertiser demand curve for advertising as a function of consumer usage and the

[100] Evans & Noel, *supra* note 1, at 669-670.

[101] *Id.*

[102] *See, e.g.,* Elena Argentesi & Lapo Filistrucchi, *Estimating Market Power in a Two-Sided Market: The Case of Newspapers,* 22 J. Appl. Econ. 1247 (2007) (analyzing network effects between readers and advertisers in the Italian magazine industry with an estimation of market power); Pierre Dubois, Adriana Hernandez-Perez & Marc Ivaldi, *The Market of Academic Journals: Empirical Evidence from Data on French Libraries,* 5 J. Eur. Econ. Ass'n 300 (April-May 2007) (linking readers and authors in the academic publishing industry); Ulrich Kaiser & Julian Wright, *Price Structure in Two-Sided Markets: Evidence from the Magazine Industry,* 24 Int'l J. Indus. Org. 1 (2006) (linking readers and advertisers in the German magazine industry); Mark Rysman, *Competition Between Networks: A Study of the Market for Yellow Pages,* 71 Rev. Econ. Stud. 483 (2004) [hereinafter Rysman, Competition] (establishing a link between readers and advertisers in the yellow pages market); Mark Rysman, *An Empirical Analysis of Payment Card Usage* 55 J. Indus. Econ. 1 (2007) (linking cardholders and merchants in the payment card industry); *see also* Elena Argentesi & Marc Ivaldi, *Market Definition in Printed Media Industry: Theory and Practice* (CEPR Discussion Paper No. 5096, 2005) [hereinafter Argentesi & Ivaldi] (summarizing antitrust cases involving market definition in (two-sided) media industries).

[103] Rysman, *Competition, supra* note 102, at 483.

quantity of advertising."[104] He finds that consumer use of a directory "increases in the directory's level of advertising."[105] At the same time, retailer demand for advertising in a directory increases "in the amount that consumers use the directory."[106] These results imply the existence of indirect network effects between the two sides.[107]

V. CASES AND AGENCY DECISIONS

Several courts and antitrust agencies have explicitly recognized the two-sided nature of certain business platforms. As noted above, in *Times Picayune Publishing,* the Supreme Court took note of the linkages between the advertiser and reader sides and the constraints each side placed on the other.[108] In *National Bancard Corp. v. Visa U.S.A., Inc.,*[109] the district court and the Eleventh Circuit relied on an analysis put forward by Professor William Baxter to recognize the two-sided nature of Visa's business platform and conclude that Visa's interchange fee was a reasonable device for balancing the demand between cardholders and merchants.[110]

> There are two possible sources of revenue in the VISA system: the cardholders and the merchants. As a practical matter, the card-issuing and merchant-signing members have a mutually dependent relationship. If the revenue produced by the cardholders is insufficient to cover the card issuers' costs, the service will be cut back or eliminated. The result would be a decline in card use and a concomitant reduction in merchant-signing banks' revenues. In short, the cardholder cannot use his card unless the merchant accepts it and the merchant cannot accept the card unless the cardholder uses one. Hence, the [interchange fee] accompanies "the coordination of other productive or distributive efforts of the parties" that is "capable of increasing the integration's efficiency and no broader than required for that purpose."[111]

The court of appeals upheld the district court's conclusion that the relevant market was "all payment devices," noting that the market was the same, whether viewed from the perspective of the plaintiff (processing agent), the defendant (credit card system operator), or consumers (cardholders).[112]

---

[104] *Id.*
[105] *Id.*
[106] Id.
[107] Id.
[108] 345 U.S. 594, 610-11 (1953).
[109] 779 F.2d 592 (11th Cir. 1986).
[110] *See id.* at 600-02.
[111] *Id.* at 602 (quoting Robert H. Bork, The Rule of Reason and the Per Se Concept, 75 YALE L.J. 373, 474 (1966)).
[112] *Id.* at 604 & n.20.

In *United States v. First Data Corp.*,[113] the U.S. Department of Justice (DOJ) sued to block a merger of First Data Corp. (First Data) and Concord EFS, Inc. (Concord).[114] First Data controlled NYCE Corporation, "the nation's third-largest PIN debit network."[115] Concord operated "STAR, the nation's largest PIN debit network" that handled "approximately half of all PIN debit transactions in the United States."[116]

Debit card systems serve consumers who use cards for payment and merchants that accept card payments.[117]

> To execute a PIN debit transaction, a customer swipes a debit card at a [point of sale (POS)] terminal and enters a PIN on a numeric keypad. After the PIN is entered, the POS terminal transmits the transaction and bank card information to a "merchant processor," which acts as a conduit between the merchant and the various PIN debit networks. The merchant processor sends the information to the appropriate PIN debit network, which switches the transaction to the issuing bank's "card processor." The card processor accesses the bank's account database to verify the PIN and ensure that the customer has sufficient funds to pay for the purchase. The card processor sends an electronic message to the PIN debit network accepting or rejecting the transaction. The PIN debit network switches this reply back to the merchant through the merchant processor to complete the transaction. The entire authorization process takes place electronically in just seconds. At the same time, the merchant acquirer "purchases" the transaction from the merchant, guaranteeing payment and facilitating settlement of the transaction.[118]

Many debit cards "can also execute 'signature' debit transactions in addition to PIN debit transactions. Signature debit transactions are authenticated like credit card transactions," except the customer signs for identification at the merchant POS instead of entering a PIN.[119]

The DOJ defined the relevant market as "the provision of PIN debit network services."[120] The DOJ explained that—compared to signature debit networks—PIN debit networks "offer a number of substantial advantages to consumers and merchants."[121]

---

113 2004 U.S. Dist. LEXIS 13590 (D.D.C. 2004).
114 United States v. First Data Corp., No. 1:03CV02169 (D.D.C. 2003) (verified complaint).
115 *Id.* ¶ 3.
116 *Id.*
117 *Id.* ¶ 16.
118 *Id.* ¶ 17.
119 *Id.* ¶ 22.
120 *Id.* ¶ 25.
121 *Id.* ¶ 23.

> PIN debit networks are generally considerably less expensive to merchants than signature debit networks, due to significantly lower interchange rates. PIN debit networks also provide a more secure method of payment than signature debit because it is much easier to forge a person's signature than to obtain an individual's PIN; consequently, fraud rates for PIN debit are substantially lower than for signature debit. Because of the increased security of PIN debit, there is no need for the complicated and expensive charge-back procedures that allow consumers to challenge signature debit transactions, thereby saving merchants additional time and money. PIN debit transactions also settle instantaneously, guaranteeing the merchant ready access to its receipts, whereas signature debit transactions usually take a day or two to settle. Finally, PIN debit networks allow for faster execution than signature debit networks. With a PIN debit transaction, customers can enter their PIN as soon as the first product is scanned. By contrast, customers cannot sign for signature debit transactions until after the entire order is totaled, prolonging the checkout process.[122]

In addition, unlike signature debit networks, PIN debit networks allow customers "to receive cash back at the register when making a purchase."[123] Finally, customers using a PIN debit network benefit from "the additional security provided by PIN verification as opposed to signature."[124]

The DOJ recognized PIN debit networks as two-sided platforms:

> The PIN debit network services market is characterized by significant network effects. Financial institutions are more likely to join networks that are accepted by many merchants. Conversely, merchants are more likely to accept networks that have many large financial institutions as members because the value of a particular PIN debit network depends in great measure on the breadth of its acceptance and use.[125]

The DOJ asserted that the hypothetical monopolist test was appropriate in a two-sided market. It explained: "There is no legal or economic support for the notion that the hypothetical monopolist test should be discarded simply because the PIN debit market is two-sided in nature . . . ."[126] It concluded that "the two-sided nature of the PIN debit market does not limit the ability of a network to profitably impose a moderate price increase on merchants."[127]

---

[122] *Id.*
[123] *Id.* ¶ 24.
[124] *Id.*
[125] *Id.* ¶ 21.
[126] United States v. First Data Corp., No. 1:03CV02169 (D.D.C. 2003) (plaintiff's pretrial brief).
[127] *Id.* § V.C.

> While the [DOJ] applied a separate SSNIP test for both merchants and issuers, it expressly accounted for the two-sided nature of payment network markets in its analysis. When the [DOJ] analyzed the impact of a SSNIP on merchants, it considered the effect of such a price increase on both merchant and card issuer behavior, including whether the SSNIP would produce a change in issuer behavior that would render the merchant price increase unprofitable. Conversely, when the [DOJ] looked at the effect of a SSNIP on issuers, it examined whether such a price increase might change merchant buying patterns to a sufficient degree to make the SSNIP unprofitable. In theory, this approach could have produced different results for each side of the market. The [DOJ's] complaint focused on the presence of a PIN debit network services market for merchants because it concluded that merchants were most likely to incur harm from the merger. The [DOJ] did not state whether it believed that PIN debit network services was also a relevant product market for issuers, but did state in its pleadings that "most banks do not perceive PIN and signature debit networks as particularly close substitutes."[128]

The government concluded that "[a] 5-10 percent increase in the fees the merchants pay for PIN debit would not change any of the above . . . . Consequently, the overwhelming majority of merchants would not reject or discourage customers from executing PIN debit transactions in response to a moderate increase in the price of the product."[129] The parties settled before trial by agreeing to divest the NYCE networks.[130]

In *GTE Media Services v. Ameritech Corp.,*[131] in assessing the competitive effects of the defendants' conduct in the Internet Yellow Pages segment, the district court recognized the two-sided nature of yellow pages businesses.[132]

> The commercial nature of the defendants' websites is revealed by the advertising revenues the defendants generate when users in . . . interact with their Internet Yellow Pages websites. Similar to publishers of the traditional, non-Internet Yellow Pages, the defendants directly earn revenue by selling advertising space to advertisers. Therefore, advertising revenue in

[128] Renata B. Hesse & Joshua H. Soven, *Defining Relevant Product Markets in Electronic Payment Network Antitrust Cases,* 73 Antitrust L. J. 709, § IV.B.1 (2006); see also Eric Emch & T. Scott Thompson, *Market Definition and Market Power in Payment Card Networks,* 5 Rev. Network Econ. 45 (2006) (discussing market definition in payment card networks markets).

[129] United States v. First Data Corp., No. 1:03CV02169 (D.D.C. 2003) (plaintiff's pretrial brief). The DOJ clarified that "consideration of the bank side of the market does not in any way suggest that the hypothetical monopolist test leads to the wrong result." *Id.* § V.C. This is perhaps best seen as an example of where the five-to-ten percent SSNIP test defines a narrow market that seems at odds with the economics of the business under consideration.

[130] United States v. First Data Corp., 2004 U.S. Dist. LEXIS 13590, at *9-10 (D.D.C. 2004)

[131] 21 F. Supp. 2d 27 (D.D.C. 1998).

[132] *Id.* at 38-39.

> the Internet Yellow Pages market substantially depends on the number of users accessing a particular website because Internet advertisers are willing to pay higher advertisement rates on websites with a higher volume of user traffic.[133]

The court concluded that the plaintiffs had adequately alleged antitrust violations, including "a conspiracy to monopolize the national Internet Yellow Pages market."[134]

> By channeling users away from competing services, such as GTE's SuperPages, the defendants not only restricted competitors from generating advertising revenue but also secured profitability for their website services. In this context, it is apparent that the purpose behind these concerted acts was to "drive . . . rival[s] from the market by exclusionary or predatory means."[135]

In defining markets in the newspaper industry, some competition authorities have focused on the characteristics of the publications' readers. In *Newspaper Publishing*,[136] the European Commission (EC) clarified that different categories of newspapers should not be considered substitutes from the point of view of buyers of advertising space because they "provide different 'channels' through which to reach different socio-economic groupings of readers."[137] The Italian antitrust authority took a similar position in *Class Editori/Sole 24 Ore*,[138] holding that advertising in "daily papers specialized in business and financial information" was a separate product market from advertising in "newspapers of general information."[139] The Italian antitrust authority concluded that "complementarities" outweighed "substitutabilities" because of "the different characteristics of readership between the two types of publications."[140]

In at least one case, in analyzing the two-sided market for printed media, the antitrust authority may have failed to "consider[] both sides of the market" when conducting econometric analysis.[141] In a case involving "two publishers or magazines and newspapers in France," the French competition authority sought to define the boundaries of the market for weekly magazines and general information and conducted an econometric analysis to ascertain whether this definition should include a larger number of titles than the definition used in previous decisions. The objec-

---

[133] *Id.*
[134] *Id.* at 42-45.
[135] *Id.* at 45 (quoting Ass'n for Intercollegiate Athletics for Women v. NCAA, 735 F.2d 577, 585 (D.C. Cir. 1984)).
[136] Case IV/M.423, Newspaper Publ'g (1994).
[137] Id. ¶ 16 (noting the "very limited substitutability between the tabloid and quality segments").
[138] Provv. n. 3336, Class Editori / Il Sole 24 Ore (1995).
[139] Argentesi & Ivaldi, *supra* note 102, at 6.
[140] *Id.*
[141] *Id,* at 8-9.

tive was to estimate the cross-elasticities between different titles in order to decide which ones were to be included in market definition.[142]

Although the authority's decision did not report its estimation results, its "very low estimates of elasticities" could have resulted from "an inadequate econometric specification."[143] The French authority's econometric model, which did not take into account "the link between readers' demand and advertising demand," has been criticized as "potentially misspecified" and having potential to "lead to biased estimations of price coefficient and related elasticities."[144]

Google's acquisition of DoubleClick, cleared after in-depth investigations by the U.S. Federal Trade Commission, the EC, and other competition authorities,[145] involved interesting issues concerning two-sided markets and is an example of a case in which focus on market definition can obscure the dynamics of the business.[146] The merger involved online advertising for web publishers. Google is an integrated advertising platform that connects advertisers to publishers: it "provides online advertising services for web publishers and advertisers in addition to operating a search-engine that sells advertising space."[147] Web publishers sell space—and the viewers who see that space—to advertisers.[148] Publishers agree to program space to accept Google ads.[149] Advertisers bid on keywords that lead Google to insert ads on web pages that have those keywords.[150] Google, acting as an intermediary, splits the ad revenue with the publisher.[151]

"DoubleClick is principally a provider of server-based software tools and services for managing online advertising for web publishers and advertisers."[152] It (1) licenses software and provides services to publishers to manage advertising on their web pages, and (2) licenses software to advertisers to manage their advertising campaigns.[153] DoubleClick uses certain efficiencies from having publishers and advertisers use the same software as a selling point.[154] Publishers use DoubleClick's software to manage advertising they sell through direct sales forces and through advertising networks.[155] They generally do not use the software with Google's ad platform, which is a complete solution.[156] Large

---

[142] Argentesi & Ivaldi, *supra* note 102, at 8.
[143] *Id.*
[144] *Id.* at 8-9.
[145] Evans & Noel, *supra* note 1, at 682.
[146] *Id.*
[147] Evans & Noel, *supra* note 1, at 682.
[148] *Id,* at 683.
[149] *Id.*
[150] *Id.*
[151] *Id.*
[152] Evans & Noel, *supra* note 1, at 682.
[153] *See* DoubleClick, *DART for Publishers,* http://www.doubleclick.com/products/dfp/index.aspx; DoubleClick, DART for Advertisers, http://www.doubleclick.com/products/dfa/index.aspx.
[154] Evans & Noel, *supra* note 1, at 684.
[155] *Id.*
[156] *Id.*

publishers will typically use DoubleClick or a competing software package for much of their ad space and will also use Google's ad platform.[157] Small publishers will often just use an integrated ad platform, such as Google's.[158]

Google's acquisition of DoubleClick involved a complicated multisided market.[159] As explained above, both Google and DoubleClick are two-sided platforms. Google operates an integrated platform; its advertising platform is the only thing that stands between the publisher and the advertiser. By contrast, DoubleClick provides an input into a decentralized platform. A publisher uses a software tool and one of several methods of intermediation, including its own sales force or advertising networks. An advertiser uses a software tool and one of several methods of intermediation, including buying agents and advertising networks.[160] DoubleClick supplies the software inputs to both sides.

Any economic effects of the Google/DoubleClick transaction result from the acquisition by an integrated platform of inputs into a decentralized platform.[161] Key issues are the extent to which the two platforms are substitutes taking both sides into account.[162] If the two types of platforms are in the same market, the inquiry should focus on the effect of the acquisition of an input into one platform type on prices and other competitive strategies.[163]

VI. CONCLUSION

Market definition helps focus antitrust analysis on a relevant set of products and businesses. It limits the amount of data gathering and provides context for consideration of possible competitive effects. The fact that one or more subjects of the inquiry are two-sided platforms does not fundamentally alter market definition analysis. However, the interdependence between the two sides of a platform and the products and businesses relevant to both sides must be considered. The standard one-sided approaches to market definition will tend to exclude the other side of the platform, placing important two-sided considerations outside of the purview of the analysis and define markets too broadly or too narrowly depending on the technique used. While these approaches can be adapted for use in cases involving multisided businesses, market definition analysis in such cases is likely to be more difficult in practice and subject to greater error than in cases involving only single-sided businesses.

---

157 For example, *Business Week* found that "[n]early all of the major Web publishers already use DoubleClick's services." *See* Robert Hof, *Google Launches DoubleClick Ad Exchange*, BUSINESS WEEK (Sept. 18, 2009).

158 For example, *USA Today* found that "[s]mall Web site operators have flocked to [Google's] AdSense as a way to attract advertising." See Jefferson Graham, *Google's AdSense a Bonanza for Some Web Sites,* USA TODAY, Mar. 10, 2005.

159 Evans & Noel, *supra* note 1, at 682.

160 *Id,* at 684.

161 Evans & Noel, *supra* note 1, at 685.

162 *Id.*

163 *See* Evans & Noel, *Industrial Organization, supra* note 1, at 682-89 (analyzing unilateral effects of the Google/DoubleClick transaction using two-sided versions of traditional econometric analyses).

CHAPTER

6 Defining Markets that Involve Two-Sided Platforms

David S. Evans and Michael D. Noel

ABSTRACT

A multi-sided platform (MSP) serves as an intermediary for two or more groups of customers who are linked by indirect network effects. Recent research has found that MSPs are significant in many industries and that some standard economic results—such as the Lerner Index—do not apply to them, in material ways, without some significant modification to take linkages between the multiple sides into account. This chapter extends several key tools used for the analysis of mergers to situations in which one or more of the suppliers are MSPs. It shows that the application of traditional tools to mergers involving MSPs results in biases the direction of which depends on the particular tool being used and other conditions. It also extends these tools to the analysis of the merger of MSPs. The techniques are illustrated with an application to an acquisition by Google in the online advertising industry.

I. INTRODUCTION

This chapter presents an empirical framework for examining market definition and unilateral effects in mergers in which one or more of the businesses that may be considered for the hypothetical market are multi-sided platforms (MSPs). MSPs provide goods or services to several distinct groups of customers who need each other in some way and who rely on the platform to intermediate transactions between them (Evans (2003a, b), Rochet and Tirole (2003, 2006)).[1] They typically reduce transaction costs and thereby permit value-creating exchanges to take place that otherwise would not occur (Evans and Schmalensee (2007a, b)). In particular they facilitate the realization of indirect network externalities, and externalities in use, between the members of distinct customer groups (Rochet and Tirole (2003)).

Many old industries are based on MSPs, ranging from village matchmakers that date from ancient times to advertising-supported newspapers introduced in the 17th century to payment cards introduced in the mid 20th century. However, an increasing number of significant modern businesses are MSPs as a result of technological changes that have drastically lowered the costs and

[1] Many multi-sided platforms have two primary sides such as advertising and readers, and much of the economic literature has focused on the case of two-sided platforms (2SPs). We will discuss it as well later in this paper.

increased the benefits of connecting diverse customer groups on a single platform. These include most internet-based businesses such as eBay, Facebook, and Google. Such businesses are creating new products and services such as social networking and are disrupting existing industries such as advertising-supported media (Evans and Schmalensee (2007b)). MSPs are likely to become an increasingly important part of merger and antitrust analysis in the years to come for several reasons.

- With the relative increase in the birth of MSPs more businesses will be based on this rather than the traditional supply-chain model in which raw materials are turned into products that are sold to wholesale and then to retail distributors.
- MSPs are prevalent in new industries that are going through consolidations based on mergers and acquisitions which must be reviewed by antitrust authorities. Recent examples include the proposed merger of Euronext and the London Stock Exchange, and Google's acquisition of DoubleClick discussed in more detail later.
- Indirect network effects tend to lead to high levels of concentration (see Evans and Schmalensee (2007a)) in industries which in turn leads to increased merger and antitrust scrutiny.
- Modern MSPs are sometimes locked in battle with other MSPs for control of an industry and are disruptive forces in traditional industries. In both cases this leads to competitors being harmed and complaints being lodged.
- MSPs have business models that are not yet well understood and engage in highly complex business strategies; unusual practices are suspect practices in our experience (Evans and Schmalensee (2007b)).[2]

[2] Indeed, some of the most visible antitrust cases around the turn of the 21st century have involved MSPs. These include the Microsoft cases in the United States (see Microsoft v United States, Supreme Court Docet No. 01-236), the European Community (see Microsoft v Commission, T-201/04), and Korea (KFTC v Microsoft, decision available at http://ftc.go.kr/data/hwp/micorsoft_case.pdf), all of which involve various aspects of the Windows software platform; the credit-card interchange fee cases in Australia(see Visa International Service Association v Reserve Bank Of Australia, N 973 of 2002), the European Union (see European Commission, "Interim Report I Payment Cards", Apr. 12, 2006), the United Kingdom (see Office of Fair Trading v MasterCard UK Members Forum Limited, CA98/05/05), the United States (see United States of America v Visa U.S.A. Inc., Visa International Corp., And MasterCard International Incorporated, 98 Civ. 7076 (BSJ)), and many other jurisdictions; mobile termination charges cases involving pricing to one side of a two-sided market that have appeared in the European Community (see Vodafone / Oskar Mobile, Case No. COMP/M.3776; Cellnet and Vodafone, Monopolies and Mergers Commission, Report to the Director General of Telecommunications, December 1998; "Decrease of the Price of Fixed-to-mobile Calls", Autorité de Regulation des Telecommunications, Press Release, 6 November 2002; Competition Commission (2003), "Vodafone, O2, Orange, and T-Mobile: Reports on References under Section 13 of the Telecommunications Act of 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks"); and the U.S. Department of Justice case involving realtor.com (see United States of America v. National Association of Realtors, Civil Action No. 05 C 5140); and actual and proposed mergers involving stock exchanges (e.g. NYSE and Euronext; Euronext and London Stock Exchange; London Stock Exchange and Germany's Deutsche Börse; London Stock Exchange and NASDAQ).

The standard tools of antitrust and merger analysis, which were developed based on the economics of single-sided businesses, do not necessarily apply in ways that are material to the analysis of competition that involves multi-sided businesses. Each side of the MSP's business influences and constrains its strategies on the other side. Antitrust analysis that focuses on one side of the business in isolation from the other side is incorrect as a matter of economics, and can lead to the wrong answer when indirect network effects are significant and are relevant for assessing the practice at issue (Evans (2003a), Wright (2004)). This chapter shows how the standard tools used for analyzing market definition and unilateral effects for mergers need to be modified when the parties are MSPs. The analysis of market definition and power has obvious extensions to other areas of antitrust.

We present an empirical framework that can be used to handle situations in which one or more MSPs may be the subject of the merger analysis. Section 2 provides an informal discussion of the analysis of market definition, market power, and unilateral effects for situations that involve MSPs and how that analysis differs from that for situations that only involve single-sided firms. Section 3 then presents our formal analysis for the special case of two-sided markets. Section 4 considers a series of examples and simulations to highlight the benefits of pursuing the correct analysis. Section 5 applies the analysis to analyzing unilateral effects for a particular example: Google's purchase of DoubleClick. Section 6 presents brief conclusions.

II. ANALYSIS OF MARKETS WITH MULTI-SIDED PLATFORMS

Consider profit maximization for a platform that serves customers groups A and B. Suppose the platform has already established prices for both groups and is considering changing them.[3] If it raises the price for members of group A, fewer A's will join. If nothing else changed, the relationship between price and the number of A's would depend on the price elasticity of demand for A's. Since, however, members of group B value the platform more if there are more A's, fewer B's will join the platform at the current price for B's. That drop-off depends on the indirect network externality which is measured by the value that B's place on A's. But with fewer B's on the platform, A's also value the platform less leading to a further drop in their demand. There is a feedback loop between the two sides. Once this is taken into account the effect of an increase in price on one side is a decrease in demand on the first side because of the direct effect of the price elasticity of demand and on both sides as a result of the indirect effects from the externalities. The change in revenue from a change in the price for A's therefore depends on the price elasticity of demand for A's and the indirect network effects between the two sides. Costs necessarily go down so long as marginal costs are positive since the number of customers has dropped on both sides. As is always the case with profit maximization, the price increase is profitable if revenues do not decline more than costs decline.

[3] To keep matters simple we consider the case where each side is charged a membership fee as in Armstrong and Wright (2007). MSPs generally involve platforms on which interactions take place customers face an access fee and a usage fee although they may choose to make some of those fees zero.

The platform would like to find the prices for each side that maximize its profits by taking these considerations into account. As Rochet and Tirole (2003) observe, one can think of these as determining the absolute and relative levels of prices. Three key results hold for two-sided platforms based only on the assumptions that there are two distinct customer groups, there are positive externalities between members of those groups, and a two-sided platform provides a good or service that facilitates exchange of value between the two customer groups in the face of these externalities:

- First, each optimal price depends on the price elasticities of demand for both sides; the nature and intensity of the indirect network effects between each side; and the marginal costs that result from changing output of each side. For the special case considered by Rochet and Tirole (2003) the profit-maximizing prices are given by:

$$p^A = \frac{\eta^A}{\eta^A + \eta^B - 1}c,\ p^B = \frac{\eta^B}{\eta^A + \eta^B - 1}c \tag{1}$$

- Second, an increase in marginal cost on one side does not necessarily result in an increase in price on that side relative to the price of the other side. From (1) we have

$$\frac{p^A}{p^B} = \frac{\eta^A}{\eta^B} \tag{2}$$

  which means the price ratio between two sides depends only on the ratio of elasticities (and not inverse elasticities), but not on marginal cost.

- Third, the profit-maximizing price for one side may be below the marginal cost of supply for that side or even negative. A common situation analyzed by Armstrong (2006) is when the platform in effect buys A's who are valued by B's.

More generally, the relationship between price and cost is complex, and the simple formulas that have been derived for single-sided markets do not apply.

Several results that are relevant for the analysis of market definition and unilateral effects follow immediately from these results. In describing these results we distinguish between mergers between "symmetric" MSPs which are defined as MSPs that serve coincident sides and "asymmetric" MSPs which are defined as MSPs that do not have at least one side in common.

The widely used Lerner Index

$$\frac{p - c}{p} = \frac{1}{\eta} \tag{3}$$

where η is the usual own-price elasticity of demand, does not accurately summarize the profit-maximizing equilibrium for MSPs when applied to a particular side. This condition does not consider the linkage between the two sides and as a result does not reflect the profit-maximizing equilibrium condition for a two-sided platform. For the special case considered by Rochet and Tirole (2003) the two-sided version of the Lerner Index is

$$\frac{p^A + p^B - c}{p^A + p^B} = \frac{1}{\eta^A + \eta^B} \tag{4}$$

Merger analyses such as the critical loss and diversion ratio that are based on the one-sided Lerner Index are therefore not correct when applied to multi-sided businesses.[4]

Merger simulation models are also misspecified when they fail to consider the multi-sided nature of the business.[5] Consider the standard conditional logit equation that underlies the basic models for differentiated-product industries:

$$\eta_{jk} = \frac{\partial s_j \, p_k}{\partial p_k \, s_j} = \begin{cases} -\alpha p_j (1 - s_j), & if\, j = k \\ \alpha p_k (1 - s_k) & otherwise \end{cases} \tag{5}$$

When applied to one side of a two-sided industry this equation fails to consider the feedback effects between the two sides—in other words that the demand by side A depends on the number of customers on side B—and that one must consider the demand by sides A and B simultaneously to properly account for all the feedback effects.

The SSNIP test for defining a relevant market does not apply without significant modifications when any of products involved in the analysis are produced by an MSP. Consider the case of a merger between two symmetric MSPs that serve the same customer groups A and B. To define the market an analyst proceeds by starting with the merger of the products that serve demand for, say, side A because that is the focus of the competition concern.[6] The set of products is expanded until a hypothetical monopolist over that set of products could raise price by, say, 5 percent or more on each of those products. That set of products then defines the market for analysis. However, by ignoring side B the analyst fails to consider that the hypothetical price increase reduces the number of side A customers available to side B, which thereby reduces the prices that side B customers will pay, and furthermore reduces the number of side B customers available to side A, which in turn reduces the prices that side A customers will pay. The link between sides A and B reduces the profitability of any price increase. The market defined by a one-sided application of the SSNIP test is necessarily drawn too narrowly from an otherwise

---

4 We discuss this in detail below.

5 See Werden, Gregory J & Froeb, Luke M (1994); Baker, J., and T. Bresnahan (1985); Feldman, R. (1994); and Nevo, A. (2000).

6 For a discussion of the SSNIP test see US Department of Justice Merger Guidelines available at http://www.usdoj.gov/atr/public/guidelines/horiz_book/hmg1.html. In particular see Section 1 on Market Definition, Measurement and Concentration.

properly conducted single sided approach because it fails to consider the tempering effects on price coming from the other side.[7]

The mistake though is more profound. The purpose of market definition is, in part, to help focus the economic analysis on a relevant but finite set of products and competitive relationships for analysis. For industries in which the multi-sided effects are sufficiently strong, market definition that excludes one of the sides of an MSP results in the failure to consider multi-sided strategies and market linkages. Failure to consider those multi-sided relationships can result in Type I and Type II errors: failing to recognize practices that may be harmful because of the two-sided relationships and condemning practices that are innocuous in a two-sided context.

The remainder of this chapter provides the toolkit for conducting merger and unilateral effects analyses for MSPs.

III. CRITICAL LOSS ANALYSIS FOR MULTI-SIDED PLATFORMS

Economists have developed a number of techniques to assess market definition and the competitive consequences of a merger following seminal contributions by Farrell and Shapiro (1990), Willig (1991), and Werden and Froeb (1994). This work and subsequent contributions build on the fundamental insight that one can infer own-price elasticities of demand from price-cost margins (the Lerner Index) and use these estimated elasticities to conduct a variety of simple analyses. Given enough data one can also evaluate the effects of a merger by estimating a demand system under some assumptions (such as differentiated-market Bertrand) about strategic interactions (see Ivaldi and Verboven (2005)).

This section uses one of the most popular parsimonious techniques – critical loss analysis – to show how these techniques can be extended and modified for mergers involving MSPs. We begin with a brief overview of one-sided critical loss analysis and then introduce the two-sided variant.

A. One-Sided Critical Loss

Critical Loss Analysis was introduced by Harris and Simons (1989) as a user-friendly implementation of the SSNIP test.[8] It compares "Critical Loss" (CL) – the percentage loss in quantity of a hypothetical monopolist's products that would be exactly enough to make an X % price increase

[7] The direction of this bias can be definitively signed when the platforms are symmetric, or alternatively when the price changes deemed significant (e.g. 5%) are symmetric across products, which is standard in market definition exercises. The bias can go in either direction when both 1) asymmetric platforms are involved and 2) asymmetric price changes are considered, as may occur in many merger analyses. Other opposing biases arise from particular calibration techniques commonly used in one-sided analyses. These will be discussed in detail later.

[8] See Harris and Simons (1989), O'Brien and Wickelgren (2003) and Katz & Shapiro (2003). See Harris and Veljanovski (2003) for discussion of the use of critical loss analysis in antitrust proceedings.

in the price of all its products unprofitable – to "Actual Loss" (AL) – the predicted percentage loss in quantity that the monopolist would suffer if it did increase prices on all its products by $X\%$. With symmetric one-sided firms, the well known formulas are given by $CL = X/(X + M)$ and $AL = X(\varepsilon^{OWN} - \varepsilon^{CROSS})$, where M is the percentage markup, and ε^{OWN} and ε^{CROSS} are the own and cross price elasticities respectively. The relevant market is found when Actual Loss equals Critical Loss for a hypothetical monopolist of the given set of products in the proposed antitrust market. If Actual Loss exceeds Critical Loss, the relevant market is expanded to include more substitutes. Otherwise, it is contracted.

Critical Loss Analysis has also been used to estimate unilateral price effects from proposed mergers. In this context, one inquires whether the merger of the parties would lead the merged firm to raise prices by $X\%$ or more. If the test shows that Actual Loss exceeds Critical Loss, the merged entity would not find an $X\%$ price increase in all its products profitable. Otherwise, it would find it profitable. This provides only a first look at whether there are significant unilateral effects and more sophisticated techniques would ordinarily be used to confirm this initial finding.

The one-sided calculations are trivial, and the technique has won widespread appeal both because of its simplicity and easy measurement of inputs.[9]

B. Two-Sided Critical Loss

We now show the one-sided formulas given above are analytically wrong when they are applied to markets involving MSPs. When an MSP increases the price to customers on one side of its business, it results not only in a loss of customers on that side but also a loss of customers on the other side of the market. This in turn causes a shift in the relative and absolute sizes of all platforms in the market, giving rise to further implications. The direction of the bias from using one-sided formulas depends upon the analytical method being used and the nature of the differences between the MSPs.

For our exposition of the biases from the use of one-sided formulas when MSPs are involved, we consider the case of two symmetric platforms serving two distinct customer groups. We return to the cases of asymmetric platforms that do not have all sides in common and MSPs with $N > 2$ in the appendix. There are two opposite biases—"Estimation Bias" and "Lerner Bias"—that occur depending on the estimation technique, and in the symmetric case the direction of the biases is unambiguous.

Estimation Bias. Suppose the analyst estimates a demand system for one of the products offered by two-sided platforms using data and following the techniques that give an unbiased estimate of the short-run own-price elasticity of demand. Here we define short-run to represent the length of time

[9] One could also calculate and compare profit totals directly across hypothetical and real worlds. The sum of firm profits before is $\pi = 2Mpq$, and the joint profit after is $\pi^{i} = 2pq(M + X)(1 + X(\varepsilon^{OWN} - \varepsilon^{CROSS}))$. We, however, maintain the Critical Loss formulation throughout the paper because of its appeal, its transparency (over black box profit comparisons), and its tangible metric – namely, how many sales a firm can afford to lose when it raises price.

it takes for customers who experience a price increase directly to respond, but before any feedback effects commence. (The feedback effects, we know, will cause additional indirect responses over time as relative and absolute platform sizes change.)[10] Since the analyst's estimate does not account for feedback effects, the full impact of the price increase on demand is underestimated. As a result, antitrust markets will be defined too narrowly, and merger analysis will overstate the increase in market power of merging parties and overstate the predicted unilateral price effects of the transaction.

Lerner Bias. An opposite bias can occur when the analyst uses observed markups to calibrate the own elasticity of demand based on the one-sided Lerner Index. This bias overstates the true short-run own-price elasticity of demand. In fact, it yields an elasticity estimate that is even larger than the true *long*-run own-price elasticity of demand. We define long-run to be as long as necessary for all meaningful feedback effects to be worked out, from the consequent changes in platform size.[11] In this case, the resulting market definition would be too broad and predicted unilateral effects of a transaction would be too small.

The situation is more complex with asymmetric platforms. When the test considers equal price changes (e.g. 5%) across all of the hypothetical monopolist's products, the biases are as above. However, if the asymmetry is significant *and* if the analyst allows for differential price changes across products, the direction of the overall bias depends on the analytical method and also the relative and absolute sizes of the price increase(s) considered. Differential price changes are more likely to be a consideration in the analysis of unilateral effects, where antitrust concern may center around particular products, rather than in market definition analyses, where the price increases deemed to be significant are generally taken as uniform (e.g. 5% or 10%) under the SSNIP test. Imagine a merger between a low-margin and high-margin (in absolute terms) platform and imagine there is concern that the prices will increase at the low margin platform only. Given properly estimated short-run elasticities, the one-sided calculation underestimates the loss on the low margin platform, but misses entirely the positive gain to the high-margin platform. If enough customers would switch to the high-margin platform, the one-sided calculation understates the unilateral price effects of the transaction. This possibility turns out to be significant in our discussion of the Google-DoubleClick merger in Section V.

Our exposition returns to the symmetric case. Call the platforms firm 1 and firm 2, and the two groups of customers A and B. Note that platform symmetry implies $P^A = P_1^A = P_2^A$ and $P^B = P_1^B = P_2^B$; we do not assume symmetry across the two sides, $P^A \neq P^B$, which are typically quite different in practice. Also, define total quantity across platforms on side s as $Q^S = Q_1^S + Q_2^S$.

[10] In practice, most likely the analyst will estimate some unknown mix of a short-run and long-run elasticity (i.e. once all feedback effects have been worked out), since indirect network effects from platform size changes are not properly accounted for. The direction of bias is the same.

[11] Our definitions of "short" and "long" are defined to be the periods in which the direct price responses and meaningful indirect price effects take place, respectively. They do not relate to calendar time and are different from the concepts of short- and long-run (+/− one or two years) generally used by antitrust practitioners.

Imagine that a hypothetical monopolist of both two-sided platforms wished to increase prices to the consumers on side A by ΔP^A and to consumers on side B by ΔP^B. Similar to the one-sided case, the gain on inframarginal sales due to the price increase on side A would be:

$$\Delta P^A (Q^A + \Delta Q^A) \tag{6}$$

However, unlike the one-sided case, the marginal loss due to the price increase on side A now consists of two components:

$$-(P^A - C^A)\Delta Q^A - (P^B - C^B)\Delta Q^B \tag{7}$$

the loss on side A *and the loss on side B* due to the price increase on A.[12] Here, marginal costs are denoted by C^A and C^B, and ΔQ^s represents the loss in total quantity on side s over the time period that is relevant for the merger review (typically one or two years). Similar equations are derived for the price increase on side B. Summing up and equating the gains to the losses, we get the Two-Sided Critical Loss Formula in response to changes in the prices on side A and B of X^A % and X^B % respectively:

$$\sum_{s=A,B} \left[R^s(X^s + M^s)\left(\frac{\Delta Q^s}{Q^s}\right) + R^s X^s \right] = 0$$

where $R^s = Q^s P^s$, the revenue earned from side *s*. Critical Loss is the set of percentage quantity reductions on side A, $\Delta Q^A / Q^A$, and side B, $\Delta Q^B / Q^B$, that would leave the hypothetical monopolist profits unchanged.[13]

The *Actual* Loss to the hypothetical MSP monopolist depends on several factors. As with one-sided firms, Actual Loss depends on the (short-run) own-price and cross-price elasticities of demand. Higher own price elasticities tend to increase Actual Loss since a price increase at a given platform results in relatively more customers switching away from that platform. On the other hand, higher (short-run) cross price elasticities between the monopolist's platforms tend to decrease Actual Loss because relatively more of the customers that switched away from one platform can be recaptured with the monopolist's other platform.

---

[12] The merger guidelines state that when determining the relevant antitrust market "the terms of sale of all other products [outside the proposed antitrust market] are held constant". In general, we would expect a hypothetical monopolist to adjust prices on side B as well as on side A. This and subsequent formulas account for both cases.

[13] A special case is when the two sides are tied together in a fixed proportion, e.g. a transaction market like credit card services, where a transaction takes place between a customer on side A (a card holder) and a customer on side B (a merchant). Suppose the proposed antitrust market is the transaction, which by definition includes both sides. Given $Q = Q^A = Q^B$ and $X^A = X^B = 5\%$, equation (8) collapses to: $|\Delta Q / Q| = X /(X + M^{AB})$ where $M^{AB} = (P^A + P^B - C^A - C^B) /(P^A + P^B))$. This is the familiar one-sided Critical Loss formula where the "product" is the transaction with a composite price of $p^A + p^B$.

Unlike one-sided firms, Actual Loss also depends on the strength of the indirect network externalities that customer groups provide to one another. Recall that an MSP is only successful because it is able to bring two distinct customer groups together in significant numbers. When an MSP increases its price on side *s* for platform *i*, there is the usual contraction in demand on side *s*, as with one-sided firms. But now because there are fewer side *s* customers, the platform is less valuable to side *r* customers. This causes a contraction on side *r* as well. The feedback effects take over, causing another contraction of s side, then r side again, and so on. The stronger the externality across groups, the greater the demand contractions will be after a price increase and the greater is the Actual Loss, all else equal.

The exact Actual Loss formula depends on the specific demand form chosen. We consider, as an example, an isoelastic demand function adapted to include the special features of MSPs.[14] Let

$$q_i^s = \alpha_s q_i^r - \delta_s q_j^r + \theta_i^s \text{ where } \theta_i^s = \mu_s - \beta_s \rho_i^s + \gamma_i^s p_s^j$$

$i \neq j$, $s \neq r$, where $q_i^s = \ln(Q_i^s)$ is the log quantity demanded by side *s* customers at platform *i*, $p_i^s = \ln(P_i^s)$ is log price to side *s* on platform *i*, and parameters α_s, β_s, and δ_s are non-negative. Since we assume competing platforms, γ_s is non-negative as well.[15] Given the log-log form, the β's and γ's are the usual short-run own and cross price elasticities for side s customers respectively. Recall that we define short-run to include the time needed for the direct response of customers impacted by the price increase but too short to include any of the indirect network feedback effects that follow. The indirect network externalities (the marginal value one side puts on the presence of the other at a platform) are captured by α_s, the within-platform cross-side externality, and δ_s, the cross-platform cross-side externality.[16] The α_s, for example, is the percentage change in quantity demanded by side *s* at platform *i* in response to a one percent change in the quantity of platform *i*'s side *r* customers. We assume $\alpha_s > \delta_s$, $\beta_s > \gamma_s$, and $\alpha_s + \delta_s < 1$ to ensure stability of the system.[17]

Solving for the reduced form equations of the q_i^s's (as a function only of prices), and taking the total derivative of each q_i^s, we derive the formula for Actual Loss to side A customers after price increases to side A and side B of X^A % and X^B % respectively:

$$L^A = \frac{(\gamma_A - \beta_A)X^A + (\alpha_A - \delta_A)(\gamma_B - \beta_B)X^B}{1 - (\alpha_A - \delta_A)(\alpha_B - \delta_B)} \tag{9}$$

[14] The results of the article carry through to other commonly used demand formulations.

[15] This demand system can be derived from a standard exogenous differentiated goods approach with an additional term included that is increasing in the consumption of consumers on the other side.

[16] The same formulas apply whether firms multi-home or single-home. While it does not change the structure of the demand system, it would be expected to impact estimates of the parameters.

[17] The first condition ensures demand on the other side of both platforms does not increase after a price increase and demand contraction at both platforms on the first side. The second ensures demand does not increase on both sides after a price increase on one side by both platforms. The third prevents exploding demand following a price increase on one side by one platform.

L^A is the percentage Actual Loss on side A at each platform and at both platforms collectively LA $= \Delta q_i^A = \Delta \ln Q^A = \Delta Q^A / Q^A$ (since platforms are symmetric), once all feedback effects have been worked out. The formula for the Actual Loss on side B is similar.

We see the two-sided Actual Loss formula for side A is a generalization of the one-sided formula. When all cross-side externalities are zero, so that each group's demands do not depend on the numbers of customers on the other side at either platform, the formulas for L^A and L^B coincide with their one-sided counterparts. When two-sided externalities are present, however, it is necessarily the case with symmetric platforms that the correct two-sided formulas give larger Actual Loss estimates than the one-sided versions, given otherwise unbiased estimates of the short-run price elasticities.

There are two reasons for the "Estimation Bias" here. First, an increase in price on side A alone causes not only a reduction in demand on side A but also a reduction in demand on side B, since the B group values a platform less when there are fewer A's. So, unlike the one-sided case, there are losses on both sides ($L^A < 0$ and $L^B < 0$).

Second, and importantly, there is a multiplier effect in two-sided industries that magnifies the immediate loss on side A and on side B over time. As discussed above, when side A contracts after its price increases, the platform is less valuable to its B customers and the B side contracts. But fewer B's decreases the value of the platform to the A customers and now the A side contracts further. Both sides contract in turn, and as a result the Actual Losses are larger than they would be in a one-sided world where absolute and relative platform size does not matter.

The multiplier effect can be seen in the denominator of the two-sided Actual Loss formulas. The denominator is necessarily less than one, so the two-sided Actual Loss formulas necessarily yield larger estimates of loss than the one-sided formulas, for a given set of parameters. The biased one-sided versions overestimate the hypothetical monopolist's ability to raise prices.

To calculate whether a given price increase or increases are likely to be profitable, one simply compares Critical Loss and Actual Loss, by substituting the values of L^s into $\Delta Q^s / Q^s$ in the Critical Loss equation (8). If the left hand side of (8) is negative, the price increase(s) will not be profitable. If positive, the price increase(s) will be profitable.

In merger analyses, the test would be performed using just the set of products or services that the newly merged firm would control. The merged entity is de facto the hypothetical monopolist and assumptions about possible cost efficiencies post-merger are easily worked into the analysis. In market definition exercises this test is repeated many times, each time expanding or contracting the proposed antitrust market, until a market is found where Critical Loss would equal Actual Loss for a hypothetical monopolist of all the products in that market. Given a market definition, the market shares of the "in" firms are calculated and then used as proxies for market power.

There are many reasons to be wary of mechanical market definition exercises such as this SSNIP test, or in relying on artificially discrete market boundaries.[18] However, for better or worse

[18] See Carlton (2007) for further discussion.

the SSNIP test is commonly used, and our purpose here is just to show that the one-sided formulas are simply wrong when MSPs are involved. The underestimation of Actual Loss means relevant antitrust markets defined on the basis one-sided formulas will be narrower than they would be under the correct two-sided calculations, and estimates of market power too high.

It is generally preferable to estimate all the parameters of the demand system whenever reliable data exists.[19] In practice, data availability is such that often parameters must be calibrated with limited information. Theoretical restrictions, like the Lerner Index that relate markups and elasticities, can be used to calibrate parameters that otherwise may be difficult or impossible to estimate. Other theoretical restrictions, like the Slutsky Symmetry rule that relates cross-price elasticities, can be used to ensure logical consistency across certain parameters.

The relationship between markups and elasticities is especially important here because markups appear in the Critical Loss formula and the elasticities appear in the Actual Loss formulas. This relationship should be consistent with economic theory.[20] In calibration exercises, it is now common practice to use the (one-sided) Lerner Index to estimate the own price elasticity from the observed markup when the latter cannot be estimated from data, perhaps because data are lacking or the analyst has insufficient time. This turns out to be a problem in two-sided settings. The one-sided Lerner Index is incorrect for MSPs and relying on it overestimates the true short-run own price elasticity of demand and overstates Actual Loss.

It is well known in the one-sided case that the percentage markup equals the inverse of the own price elasticity at the profit maximizing output. In the two-sided case, we can derive the first order conditions for profit maximization. Each platform *i* chooses prices on each side to maximize:

$$\max_{p_i^A, p_i^B} \Pi_i = \sum_{s=A,B} q_i^s(\vec{p})^* (p_i^s - c_i^s) \tag{10}$$

[19] There are several empirical papers that estimate two-sided demand systems in particular industries in order to establish the existence of indirect network effects. Although generally not in the context of merger or market definition analysis, these highlight the importance and challenges of proper demand identification in a two-sided context. For example, see Rysman (2004) establishing a link between readers and advertisers in the yellow pages market, Rysman (2007) for cardholders and merchants in the payment card industry, Argentesi and Filistrucchi (2005) for readers and advertisers in the Italian magazine industry with an estimation of market power, Wright (2004) and Kaiser and Wright (2006) for linking readers and advertisers in the German magazine industry and Dubois, Hernandez-Perez, and Ivaldi (2007) linking readers and authors in the academic publishing industry. Argentesi and Ivaldi (2005) recap antitrust cases involving market definition in (two-sided) media industries. Using limited data for the French magazine market, they include an estimation of subscriber demand elasticities using the price of advertising as an instrument, suggesting a two-sided linkage. Elasticity estimates differ with and without the instrument.

[20] Some authors disregard the Lerner index as too simple and unrealistic for application in specific real-world antitrust cases (see Scheffman and Simons (2003) and Harris (2003)). We agree with Katz and Shapiro (2003) that there should be a presumption of the theoretical markup-elasticity relationship pending evidence to the contrary. It is possible, however, to estimate both markups and elasticities independently of one another directly from the data and sidestep these issues.

The two first order conditions simplify to:

$$R_i^s + R_i^s M_i^s \varepsilon_{ii}^{ss} + R_i^r M_i^r \varepsilon_{ii}^{sr} = 0 \tag{11}$$

for s = A,B, where:

$$\varepsilon_{ii}^{ss} = \frac{(\alpha_A\delta_B + \alpha_B\delta_A)\gamma_s + (1-\alpha_A\alpha_B - \delta_A\delta_B)\beta_s}{(1-(\alpha_A - \delta_A)(\alpha_B - \delta_B))(1-(\alpha_A + \delta_A)(\alpha_B + \delta_B))}$$

$$\varepsilon_{ii}^{sr} = \frac{(\alpha_A\delta_B + \alpha_B\delta_A)(\alpha_r\gamma_s + \beta_s\delta_r) + (1-\alpha_A\alpha_B - \delta_A\delta_B)(\alpha_r\beta_s + \delta_r\gamma_s)}{(1-(\alpha_A - \delta_A)(\alpha_B - \delta_B))(1-(\alpha_A + \delta_A)(\alpha_B + \delta_B))}$$

We call ε_{ii}^{ss} and ε_{ii}^{sr} the long-run own-price elasticity and the long-run cross-side price elasticity, in absolute value, once all feedback effects have been worked out. The i subscripts on the ε's indicate these are elasticities with respect to changes in the price of the side s good *by firm i only*.

The first order conditions do not reproduce the one-sided Lerner Index formula. They do not even uniquely identify the values of the own price elasticities from the markups, but rather only constrain the relationship between the own and cross price elasticities and the indirect network externalities.

To see the "Lerner Bias" most easily, imagine for a moment symmetry across sides $\alpha_A = \alpha_B$ and $\delta_A = \delta_B$. The markup equation for side s then simplifies to $M_i^s = (\varepsilon_{ii}^{ss} + \varepsilon_{ii}^{sr})^{-1}$. Now note that the long-run own price elasticity, ε_{ii}^{ss}, is necessarily larger than β (the usual short-run own price elasticity), due to the indirect network effects. The long-run cross-side price elasticity, ε_{ii}^{sr}, which is zero in a truly one-sided industry, is greater than zero in absolute value here.[21] Consequently, for a given set of elasticity estimates, equilibrium markups are lower that the one-sided Lerner Index would imply. The converse is that for a given set of markups observed in the data, the true short-run own-price elasticities of demand are lower than the one-sided Lerner Index would suggest. The reason is that the indirect network externalities penalize price increases more, so the short-run own price elasticity must be especially low in order to support markups at a given level.

The bias is strong − since $\varepsilon_{ii}^{ss} > 0$, the one-sided estimate is even larger than the true *long-run* own-price elasticity, and this results in a bias in the opposite direction. Market definitions will be set too large, or the expected price increases from a merger will be assumed too small.

The comparative statics of the two-sided markup equations give insight into the problem. As with the one-sided calculation, a higher short-run own elasticity β (which increased both

[21] Recall this is the long run change in quantity on side r at platform i due to a change in the price of side s at platform i, once all feedback effects have been worked out. There is no direct effect of side s prices on side *r* demand, but there is an indirect effect of side s prices on side r demand for two-sided platforms, via changes in side *s* demand.

ε_{ii}^{ss} and ε_{ii}^{sr}) lowers markups. But unlike the one-sided case, a higher short-run cross price elasticity of demand γ also results in lower markups (again increasing ε_{ii}^{ss} and ε_{ii}^{sr}). The reason is that a higher cross price elasticity causes a higher proportion of the side A customers who switch after a side A price increase to buy from platform *j* rather than buy nothing instead. Since platforms are two-sided, the relative value of firm *i*'s platform to the side B customers falls when more A's move to *j*, causing a demand contraction on the B side as well. The feedback effects begin, and platform *i* contracts further.

Greater indirect network externalities also work to decrease markups. First, by raising price a platform triggers the market shrinking feedback effect causing its demand to repeatedly contract on each side. It is easiest to see this by noting $M_i^s = (1 - \alpha)/\beta$ when $\delta = 0$ in the symmetric case. A greater cross-side own-platform externality α penalizes price increases more and acts to lower equilibrium markups.

Second, and surprisingly, a higher cross-side cross-platform externality δ also acts to lower markups. To see this, note that $M_i^s = (1 - (\alpha + \delta))/\beta$ as $\gamma \rightarrow \beta$ in the limit in the symmetric case. The parameters α and δ now appear as a sum, rather than a difference (as they did in the Actual Loss formulas). Only platform *i* is changing price, not the hypothetical monopolist of both products. A higher δ means that in response to an increase in the price to side A at platform *i*, more side B customers switch to platform *j* rather than to nothing at all. This makes *j* more attractive to side A, causing the market shrinking feedback effects to begin, and hurting platform *i*. Thus, higher values of both α *and* δ penalize price increases more and lead to lower markups in equilibrium.

Overall, for a given set of short-run own-price and cross-price elasticities, two-sided pressures lower equilibrium markups below that which the one-sided Lerner Index would predict. One-sided estimates of short-run own-price elasticity derived from markups overstate the true elasticity and understate the profitability of potential price effects. We now turn to a demonstration of the potential "Estimation" and "Lerner" Biases from using one-sided formulas in situations involving symmetric MSPs.

IV. COMPARISON OF ONE AND TWO-SIDED CRITICAL LOSS: SOME SIMULATIONS

Assume that two platforms each serve sides A and B, and as an example, assume a relevant antitrust market is proposed that would include the products on side A of these two platforms. The test would be the same if we were considering a merger between the two platforms and wanted to know if a given price increase on side A would be profitable for the merged entity.[22] We again maintain the assumption of symmetric platforms.

[22] We could also assume price increases on both sides A and B, and the same qualitative results still hold.

It is useful to rearrange the Critical Loss equation (8), substituting in L^A and L^B for $\Delta Q^A / Q^A$ and $\Delta Q^B / Q^B$, to get:

$$\omega \frac{X^A + M^A}{X^A} L^A + (1 - \omega) \frac{X^B + M^B}{X^B} L^B = -1 \tag{12}$$

which holds when Critical Loss equals Actual Loss exactly. It is clear that the equation is really a weighted sum of the ratios of Actual Loss to Critical Loss on each side, where the weight is given by

$$\omega = \frac{R^A X^A}{R^A X^A + R^B X^B}$$

We refer to the left hand side of this equation, in absolute value, as the Actual-Critical Loss Ratio (ACR). If the ACR is greater than 1, Actual Loss exceeds Critical Loss, and the price increase would not be profitable. The relevant market would be expanded. If ACR is less than 1, the price increase is more than profitable and the relevant market contracted. The ACR represents a measure of the closeness of the Actual and Critical Loss.[23] Note that if the industry were truly one-sided, so that LB = 0 after a price increase on side A, this equation reduces to the ratio of Actual Loss to Critical Loss that would be applicable with one-sided firms.

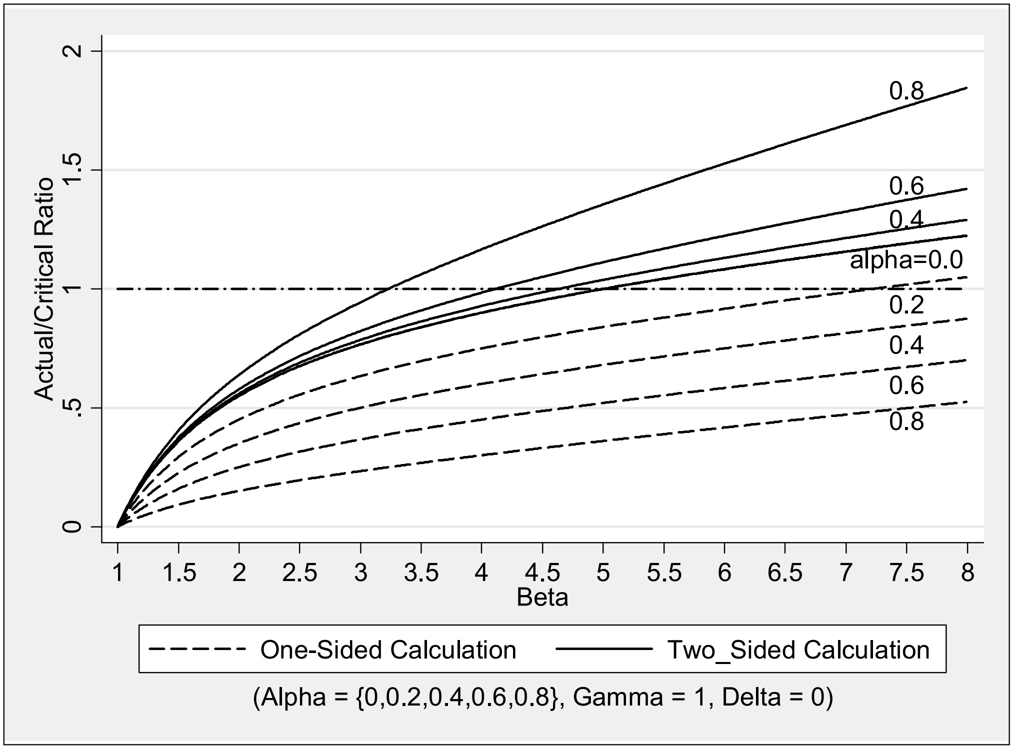
To begin, assume that the short-run own-price elasticities are well estimated and the analyst does not depend on the (one-sided) Lerner Index. We see the one-sided calculations are biased towards a lower ACR, and the bias grows as the indirect network externalities become more important. Figure 1 plots the ACR against the own-price elasticity β for several different values of the cross-side own-platform externality α. Recall the parameter α represents the short run percentage change in a platform's side A business due to a 1% change in demand on its side B business, and vice versa. The figure includes both the ACR calculated using the one-sided formula (which assumes α and δ are both zero), and also using the correct two-sided formula for MSPs. Other parameters are held constant, and the price increase to side A is taken to be 5%.[24]

First, and as we would expect, the ACR increases as the short-run own-price elasticity increases in both calculations for a given α. There is an indirect effect of higher β through markups that increases Critical Loss, but the direct effect of β that increases Actual Loss dominates, and the ACR rises with β.

The calculated effect of the own-platform cross-side externality α on the ACR, however, depends on the calculation being used. In the two-sided calculation, a higher α is associated with stronger feedback effects, which result in greater Actual Loss and greater ACR following a price

[23] The derivation assumes $X^A \neq 0$ and $X^B \neq 0$. When $X^B = 0$, the second term reduces to $(R^B M^B / R^A X^A) L^B$.
[24] The assumption of 5% is often used but arbitrary.

FIGURE 1 Actual/Critical Loss Ratios, One- and Two-Sided Calculations

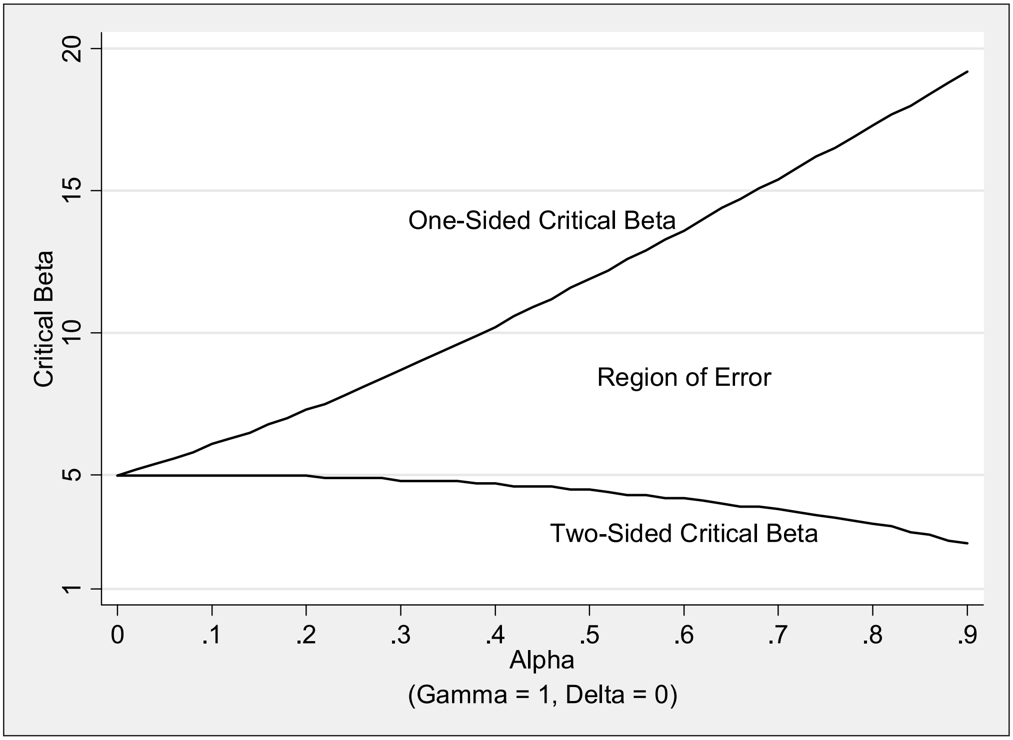


increase, all else equal. In contrast, the ACR falls under the one-sided calculation. This is because the Actual Loss calculated with the one-sided formulas $((\beta_A - \gamma_A) X_A)$ is independent of α, but the Critical Loss is greater since observed equilibrium markups are lower with higher α. The gap between the ACRs under the two calculations grow with higher α, and with higher β.

Figure 2 plots the "critical beta" under each calculation against the cross-side own-platform externality α. The critical beta is the value of the own price elasticity, conditional on the other parameters, that would cause Actual Loss to exactly equal Critical Loss. The area between the one and two-sided critical β's is the region of error: the combinations of α and β for which a contraction of the proposed antitrust market would be called for under the one-sided calculation when an expansion of the proposed market should be called instead. In merger analyses these would be the combinations of α and β for which the analyst would conclude from the one-sided calculation that the hypothetical monopolist in a market definition analysis or the merged entity in the unilateral effects analysis would significantly increase prices when in fact it could not.

The figure shows that as the indirect network effects become stronger, the region of error grows wider. The corresponding figure for the "critical gamma", with β held fixed (not shown), yields the same conclusion.

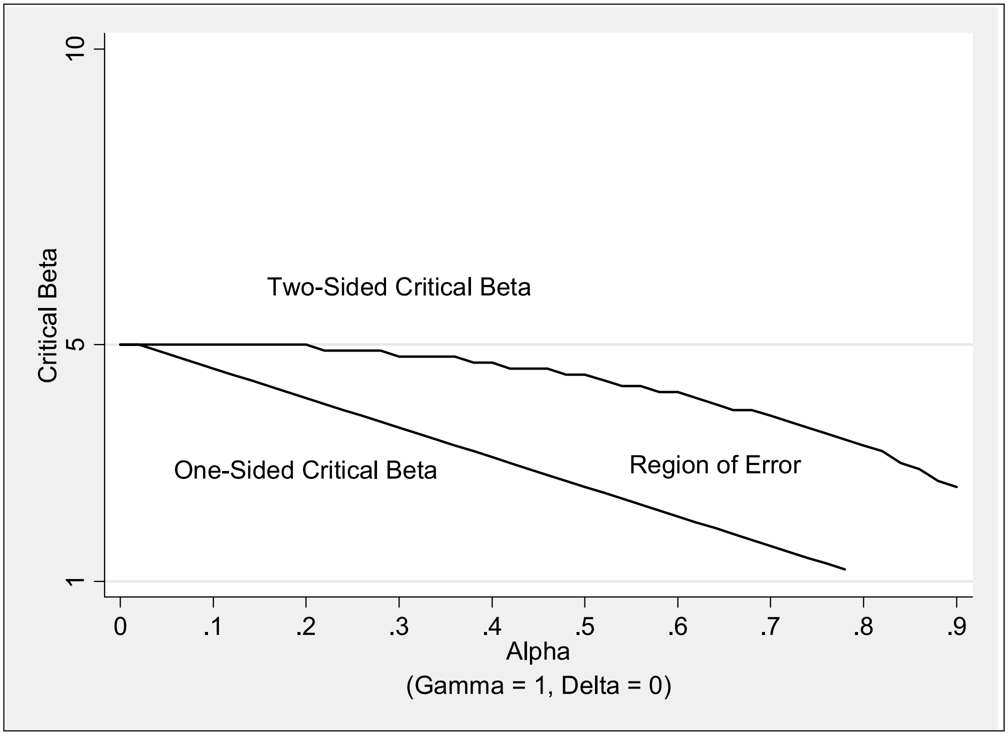
FIGURE 2 Region of Error, One- and Two-Sided Calculations



In these examples, it is assumed that the short-run elasticities are properly estimated; only the formulas used differ. However, if the analyst relies on the (one-sided) Lerner Index to estimate the own price elasticity of demand, and then uses the usual one-sided formulas, the bias goes in the other direction. The analyst will overestimate the short-run and long-run own price elasticities of the MSPs and will overestimate Actual Loss. The one-sided Critical Loss calculation is unchanged (because it is based on observed markups) and so the ACR will be erroneously high. In other words, profitable price increases will not be expected for even relatively low values of the true short-run own-price elasticity β. The critical betas and the region of error are plotted in Figure 3, and this time the critical betas are lower under the one-sided calculation than under the two-sided one. In the region of error, the analyst who relies on the one-sided calculations would recommend an expansion of the antitrust market when a contraction should be called for, or would conclude that merging platforms could not profitably raise prices after the merger when it fact they could.

The asymmetric platform case, given by formulas set out in the appendix, is more involved. When platforms are asymmetric but the price increases considered on any side are equal across platforms, the biases move in the same directions as in the symmetric case. However, when platforms are asymmetric *and* price increases considered are also allowed to differ across products,

FIGURE 3 Region of Error, Mismeasured β in One-Sided Calculations



the direction of the bias also depends on the degree of asymmetry and the particular mix of price increases. Differential price increases are most likely of interest in merger analyses where antitrust concern may center around particular product offerings. (In market definition exercises, the price increases deemed significant is generally held equal across platform (e.g. 5% or 10%).) The reason differential price increases are important is because they result in a shift in the relative sizes of the platforms and generate feedback effects. If the platform is asymmetric enough, the impact of certain price increases will be overstated, and others understated.[25]

As an example, consider a proposed merger where the platforms differ significantly in initial profitability. A price increase at the low-margin platform will be beneficial for the firm if network effects are strong and doing so generates enough customers switching to its high-margin platform.

[25] Across all four potential price changes of a given equal amount, the "average" bias from those price changes collectively goes in the same direction as that in the symmetric case, which in turn depends on the analytical method used. However, in the asymmetric case there is variation in the size of these biases and with sufficiently asymmetric firms, some may have the reverse sign. Because differential price changes act as weights in Actual Loss (equation A.4), the overall bias can go in the opposite direction of the symmetric case if the reverse-signed biases are sufficiently highly weighted.

One-sided calculations do not account for this possibility, and as a result, understate the profitability of this price increase. If we consider a price change at only the high-margin platform, we can get an opposite bias. The profitability of the price increase would be overstated by a one-sided calculation as long as the gain from expansion of the low-margin platform cannot compensate for the contraction of demand at the high-margin platform. Thus, when asymmetric platforms and asymmetric price changes are considered, the direction of the biases depends on the specific question being asked as well as the method used.

V. ANALYSIS OF UNILATERAL EFFECTS: GOOGLE'S ACQUISITION OF DOUBLECLICK

On April 13 Google announced that it had reached an agreement to purchase DoubleClick for $3.1 billion. Google is a provider of online advertising services for web publishers and advertisers in addition to operating a search-engine that sells advertising space. DoubleClick is principally a provider of server-based software tools and services for managing online advertising for web publishers and advertisers.[26] This section uses data on margins for Google and DoubleClick along with other assumptions on their competitive relationships to show how the techniques used above can be employed. Our purpose is not to assess whether or not the merger would in fact have competitive effects – that would entail a much more detailed analysis than is possible here – but rather to illustrate a practical application of the two-sided critical loss framework.[27] Furthermore, there is a remarkable lack of publicly available data on the relevant products and services and many of the numbers reported below are based on estimates from a variety of sources.

Web publishers make space available on their web pages for advertisements. Their ad inventory corresponds roughly to the number of people that will view each space over some given time period.[28] Advertisers are on the other side of the market. They buy advertising inventory to reach consumers. The online advertising business that is relevant to the Google-DoubleClick transaction is basically about how this advertising inventory is bought and sold.[29]

[26] As of the writing of this paper the merger is under review by the Federal Trade Commission and the European Commission.

[27] The authors have been consultants to several parties that have sought close scrutiny of the transaction on the part of the relevant competition authorities and have argued that the transaction could significantly lessen competition in various markets.

[28] Advertisers value the demographic and other characteristics of viewers in addition to their sheer number. Thus views by high-income residents of Boston are more valuable to the local BMW dealer than views by low-income resident or high-income residents of Topeka.

[29] The more developed part of the online advertising business consists of search. Search-engines supply advertising inventory on their search-results pages and thereby compete with regular publishers for advertiser dollars. Search engines are also relevant to Web 2.0 because they are the way people often find websites either through organic search results or through advertising.

Publishers create advertising inventory by designing their web pages to accept graphical, text, or video ads in various portions of the page. The amount of inventory they can supply then depends on the viewers they attract to these pages. This supply is highly heterogeneous for two primary reasons. First, like newspapers, some space is seen as more desirable than others – the top right is more desirable than the bottom left because people are more likely to pay attention to the former than the latter. Second, some viewers are worth more to advertisers than other viewers, and the technology of online advertising enables publishers and advertisers to establish prices for viewers with particular characteristics.

Like all sellers and buyers, publishers and advertisers require ways to identify optimal trading opportunities and to establish transaction prices. There are two major ways that this "intermediation" occurs. First, it can occur *directly* through bilateral exchanges between publishers and advertisers. eBay for example, may sell Nokia the right to present an advertisement in a particular spot to viewers with specific characteristics by having its sales agents deal directly with Nokia's buying agents. Second, it can occur *indirectly* through multilateral exchanges between publishers and advertisers using advertising networks. Hearst Publishing may sell ValueClick – an advertising network – advertising inventory from its various online newspaper and magazine properties, which ValueClick will then sell to advertisers who want to reach the kind of viewers that Heart Magazines has. The advertiser in this case typically buys access to a type of viewer –"fashion conscious young women in upscale locations" – but has not specifically bought space on *Cosmopolitan's* website. The extent to which advertisers and publishers use direct and indirect methods of distribution for advertising varies. Smaller ones typically rely only on indirect methods because it is not economical to carry the cost of salespeople and purchasing agents. Google's AdSense network is especially popular with small advertisers and publishers in part for this reason.

This section concentrates on large web publishers, which account for the preponderance of advertising revenue and large advertisers, which account for the preponderance of online spending.[30] A significant part of the advertising inventory bought and sold by these large advertisers and publishers involves bilateral exchange. Large publishers either have direct sales forces or hire third-party sales reps to sell their ad inventory. Likewise, large advertisers have purchasing agents or, more often, use media buyers at their advertising agency to purchase ad inventory. Advertising inventory sold this way is said to be "reserved". Large publishers often sell their "premium ad space" this way.

Large publishers often also rely on other intermediaries to sell ad inventory indirectly that they have not "reserved" for advertisers directly. These intermediaries are called "advertising networks". Publishers may use ad networks because they are more efficient than a direct sales force for some, or all, of their ad inventory; or because they have excess inventory that they have not sold directly,

[30] A large publisher refers to one which is sufficiently large to use stand-alone tools to serve their directly sold ads and any remnant ads sold through non-integrated and/or integrated ad networks. This includes at least the top 500 publishers, and accounts for the majority of advertising revenues. Smaller publishers do not find it economically efficient to hire direct sales forces. They usually rely on an ad network to sell their ad inventory.

perhaps because of spikes in viewers. Advertisers use ad networks because it is another way to reach viewers. Hard estimates are difficult to come by, but advertisers and publishers we have talked to seem to agree that somewhere around 30 percent of advertising revenue for large publishers is sold indirectly, as well as more than 50 percent of advertising impressions (the number of viewers of an ad); this figure appears to vary significantly across publishers depending on the way they have chosen to manage their advertising business. On average, advertising inventory sold indirectly costs less than advertising inventory sold directly because the space is less desirable than the directly sold space.

There are several aspects of advertisers' management of their online campaigns, as well as publishers' management of their sales of advertising inventory. As discussed previously, advertisers and publishers need intermediation services. For bilateral exchanges they may use some combination of in-house and third-party providers, while for multilateral exchanges they mainly use advertising networks. Advertisers and publishers likewise need management, reporting, and technology solutions such as those offered by DoubleClick and aQuantive. These tend to be server-based software that can help manage advertising inventory and campaigns that may involve millions of ad impressions (that is, views of an ad by an individual) a day. These server-based software tools are highly sophisticated mission-critical applications.

Large publishers usually use a publisher tool such as DoubleClick's Dart for Publishers (DFP).[31,32] This tool is typically hosted on a web server maintained by the provider. The publisher hardcodes links to the publisher tool to fill the ad space for which it wants to use the management, reporting, and serving capabilities of the publisher tool. It will also typically integrate the publisher tool into many other aspects of the website technology and business practices. As an example of how such publisher tools are used, consider a visit to www.cnn.com. When the entertainment page of www.cnn.com is clicked, a decision is made as to which ad to present to the user and, once chosen, the ad is displayed so the user can see it, and the publisher can earn some money. The publisher – in this case www.cnn.com – uses its publisher tools to check whether the particular ad space that the user is about to see has been "reserved", and if not, whether there is an ad network that can fill that inventory space. The publisher tool then retrieves the chosen ad and presents it to the user. This entire process happens in the blink of an eye. Publishers will typically only use one publisher tool (i.e., they "single-home" on tools).

Large advertisers and advertising agencies often have an advertiser tool such as DoubleClick's Dart for Advertisers (DFA).[33] Large advertisers typically manage advertising on hundreds of websites and across numerous products using many methods of online advertising. This tool helps them manage these various campaigns. Advertisers usually use one advertiser tool although advertising agencies may use several. However, single-homing does appear to be the norm.

[31] See http://www.doubleclick.com/products/dfp/index.aspx

[32] A handful of mega-large publishers such as MSN have their own proprietary tools but most others use a third-party tool.

[33] See http://www.doubleclick.com/products/dfa/index.aspx

A few providers offer more or less complete solutions for advertisers and publishers. Google's AdSense/AdWords platform is one of these. Publishers can hardcode ad space to Google's AdSense[34] which takes care of everything – selling the inventory, managing the ad space, serving ads to the viewer, and sending the publisher a portion of the proceeds after taking a commission.[35] Advertisers, likewise, can buy space from the Google Content Network[36] through AdWords[37] (which bundles Google's search-based and contextual-based advertising products). Yahoo! and Microsoft offer similar all-inclusive solutions. These solutions have all resulted from leveraging the technologies developed for search-based advertising – especially the keyword bidding auctions – to the buying and selling of publisher ad inventory. Some other ad networks are also integrated to lesser degrees; they may offer publishers serving technologies so that publishers can hardcode ad network into particular space. Many large publishers use an integrated platform for contextual ads and an unintegrated platform (based on a particular publisher tool) for non-contextual ads; the unintegrated platform is used to access multiple standalone ad networks, as well as ads sold directly.

Online advertising is highly differentiated. It comes in different forms such as graphic ads that mix pictures and text, all text ads, and video ads; appears in different places on websites; and is targeted to viewers in various ways. It is also bought and sold in different ways which leads to differentiation in terms of the distribution channels used by advertisers and publishers. The underlying economics and driver of substitution are fairly straightforward. Publishers are interested in maximizing the rate of return on investment in their ad inventory. Advertisers are interested in maximizing the rate of return on investment from their advertising spending- which ultimately means getting consumers to buy things. Advertiser and publisher decisions on ROI are interlinked because online advertising is fundamentally a two-sided business, as discussed below.

Surveys of large websites show that DFP is used by somewhat more than half of these websites. Larger sites are more likely than smaller sites to have DFP. DFP has about 63 percent share based on page views.[38] DFA holds roughly a 26 percent share of non-search advertiser tools.[39] Google is

[34] See https://www.google.com/adsense/login/en_US/

[35] In the online advertising Traffic Acquisitions Costs (TAC) refers to what an advertising platform pays for traffic. Google pays TAC to publishers in return for contributing their advertising inventory to the Google Content Network. The publisher receives TAC. In this case 1-TAC is the percent commission paid by the publisher to Google for selling its ad inventory. If Google pays the publisher 80% of the revenue that Google receives from the advertiser the publisher has paid a commission of 20% to Google.

[36] See https://adwords.google.com/select/afc.html

[37] See http://adwords.google.com/select/Login

[38] The survey conducted by LECG shows that AdSense is used on 50% of all websites sampled. DoubleClick is used on 63% of all websites sampled. A second survey conducted by Keystone Strategy for American viewers shows that AdSense is used on 45% of surveyed sites while DoubleClick is used on 61%.

[39] This is based on aQuantive's estimates of shares for itself and DoubleClick in advertiser tools (among providers that sell those as standalone non-integrated products), adjusted for those advertiser tools that are sold on an integrated basis (with the ads), by providers such as AdSense.

the leading integrated ad platform. Estimates suggest that Google accounts for 51 percent of the ad revenue through the indirect channel and 27 percent overall. The unintegrated ad networks account for an estimated 45 percent of ad revenue through the indirect channel and 25 percent of web publisher ad spending overall.[40]

Following the acquisition, Google would have control of a key input into the unintegrated channel for selling ads directly and indirectly. It would have the ability to coordinate the price of DFP, and through DFP, the overall price of distribution through the direct sales and through the ad networks.[41] The question we consider is whether the combined entity would have the incentive to alter prices significantly. Two-sided effects are important for analyzing this question because changes in the prices of the inputs alter the advertiser's demand for space as well as the publisher's demand for selling through a particular method.

Table 1 reports the data and sources we have relied on. The key data is that on margin, which shows that both Google and DoubleClick earn roughly 75 percent gross margins. Ad inventory sells on average, with great dispersion, for about $2 per thousand impressions (that is, views by users). Google sells on a cost per click basis but obtains the equivalent of approximately 40¢ per thousand impressions based on the evidence we have seen.[42] Publishers typically pay 40¢ per thousand impressions on average to the standalone ad networks and 5¢ per thousand impressions to the publisher tool provider.

We want to ask whether the merged Google-DoubleClick entity would have the ability to increase the price of DFP by X%, holding all else constant. This was a central point of concern with the merger. Google's AdSense is an integrated product, and so charges higher markups than DoubleClick's DFP. This suggests it might be advantageous for Google to raise the price of DFP and shift publishers to its AdSense network.[43] In the analysis, we take X to be either 5% or 10%. Since the platforms are inherently asymmetric, the analysis uses the formulas set out in the appendix. Ex ante, the direction of bias from using the one-sided calculation in the two-sided setting is unknown.

[40] Source: Keystone Strategy.

[41] We focus on price effects as is traditional in this sort of analysis. Google could engage in a variety of other strategies post-acquisition that could involve the exercise of acquired market power but that do not necessarily involve increasing prices.

[42] The revenue paid by advertisers by publishers is estimated at around $2 CPM of which publishers keep approximately 80%. See http://www.mydigitallife.info/2006/10/21/google-adsense-giving-publishers-average-of-78-revenue-share/ and http://www.webmasterworld.com/forum89/6913.htm

[43] In the two-sided calculation, we attribute 2/5th of the Google unit price and revenues to the publishing side and 3/5th to the advertising side, to match the DoubleClick ratio. These are the expected side-specific prices that Google would charge had advertisers paid publishers directly for the advertising space and each paid Google a commission for the value of the tools and intermediation services provided. This avoids the need for negative prices and negative externality parameters in the calibration, which adds transparency.

TABLE 1 Data and Sources

	DFP	DFA	AdSense	Source
Unit price (cpm)	5¢	7.5¢	40¢	Discussions with publishers. The revenue paid to Google by advertisers is estimated at around $2 CPM of which publishers keep approximately 80%.
Percentage gross margin	75%	75%	75%	DoubleClick's gross margin according to its 2004 Form 10K was 74%. According to Google's 2006 Form 10K its costs of earning revenue (net of traffic acquisition costs paid to publishers) are around 10% of total revenues. This implies its gross margin on AdSense is 90%.Its overall gross margin (for search and non-search ad programs) has been between 50-70% in the past five years. We use 75% as reasonable.
Revenues	$325m		$196m	Google's 2004 revenues on third party ads after paying publishers. DoubleClick's 2004 revenues for "Tech Solutions" which appears to mainly be comprised of DART products.

Given the limited data available, calibration is the most feasible approach.[44] For the one-sided calculation, there are four parameters to estimate (all on the publisher side), using a combination of data and restrictions based in economic theory. First, we use the theoretical Lerner Index to fix the own-price elasticities given the observed markups (two restrictions). As discussed, the use of the one-sided Lerner Index in a two-sided setting yields a biased estimate of the true price elasticities. We also assume Slutsky symmetry, which equates quantity-weighted cross-price elasticities across platforms (one restriction). We then just need to calibrate one of the cross-price elasticities and we could calculate Critical Loss and Actual Loss. However, instead we employ the common and useful technique in calibration exercises to set Critical Loss equal Actual Loss (the fourth restriction) and then back out the values of the cross-price elasticities (connected together by Slutsky symmetry) that would exactly equate CL and AL.

[44] Given enough time a competition authority could possible obtain sufficient data from members of the industry and through surveys to obtain estimates from demand-system estimation.

In the two-sided model there are sixteen parameters total in the demand system that we need to estimate using a combination of data and economic theory, i.e. sixteen restrictions. First, we use the two-sided version of the Lerner Index given by equations (11). Conditional on the externality parameters, this yields a linear relation between each own-price elasticity and its corresponding cross-price elasticity. We keep Slutsky symmetry, a similar symmetry for the externality parameters, and seek out the "most symmetric" solution. We assume two-sided effects are conservatively small, with own-platform cross-side externalities of 10%. The restrictions are explained in more detail in the footnote.[45]

In total, these add to fifteen restrictions. Again, we equate Critical Loss and Actual Loss as the sixteenth restriction, and back out the own- and cross-price elasticities (connected together by the other restrictions) that would exactly equate CL and AL.

Following common practice, we report the easily interpretable "critical diversion ratios" and "critical switching levels" rather than the β's and γ's themselves.[46] The critical diversion ratio is defined as the fraction of the quantities lost at DoubleClick's DFP that needs to be immediately recaptured by Google's AdSense platform, for a price increase in DFP to be profitable once the indirect network feedback effects work themselves out. It is effectively the quantity-weighted ratio of the cross-price elasticity (of AdSense's quantity with respect to the price of DFP) to the own-price elasticity (of DFP). The Critical Switching Level is defined as the fraction of the total quantities of DFP that needs to be immediately recaptured by AdSense after a price increase for the increase to be profitable once all feedback effects are worked out. It is calculated as the AdSense-quantity-weighted cross-price elasticity to total DFP quantity.

[45] The two-sided Lerner Index accounts for four restrictions, each relating an own-price elasticity with a corresponding cross-price elasticity. We assume Slutsky symmetry that ensures consistency in the cross-price elasticities on each side (four restrictions) and a similar restriction to ensure consistency across the indirect network externalities on a given side (four restrictions). Our results do not change qualitatively if we loosen the assumption of Slutsky symmetry, however. We assume Google's own-platform cross-side externality is 0.1 on each side, and its cross-side cross-platform externality 0.05 on each side (four more restrictions). Under Slutsky symmetry in the two-sided parameters, the corresponding values for DoubleClick are 0.1 and 0.005. We also assume the "most symmetric solution" across sides (one restriction). This is because, conditional on the externalities, many combinations of β and γ are technically possible. A relatively low β on one side could exist and satisfy the markup condition if the corresponding γ is extremely high. Note that the effect of a change in β is about $1/\alpha$ times the effect of a change in γ in the markup equations. But this causes an opposite stress on the other side, in order to equate AL and CL (see below). This tends to force the other β to be relatively low, and the other γ to be extremely small – generally negative with a similarly large magnitude as the γ on the first side. The more reasonable solution is that which yields the most symmetric β's and γ's, conditional on all the other constraints. In a fully symmetric model this ensures a symmetric solution; in an asymmetric model the β's and γ's still differ across sides. It is implemented by minimizing the sum of squares of the γ's across sides (alternatively, the sum of squares of the β's across sides.) The total number of calibrated values plus restrictions adds to 15. We could then fix one of the β's or one of the γ's to complete the model. Instead, we force CL=AL as the sixteenth restriction, and back out the critical β's and γ's (tied together by the above constraints) that would result in CL equaling AL.

[46] See Shapiro (1996).

TABLE 2 Critical Diversion Ratios and Switching Levels

	5% Critical Price Increase		10% Critical Price Increase	
	One-sided Calculation	Two-sided Calculation	One-sided Calculation	Two-sided Calculation
Critical diversion ratios	0.83%	0.00%	1.67%	0.01%
Critical switching levels	0.06%	0.00%	0.22%	0.01%

Table 2 reports the critical diversion ratios and critical switching levels under our best data assumptions, for both the one-sided and two-sided calculations, and for critical price increases (X) of 5% and 10%.

Given the high markups in the industry, and the large differential in absolute margins between AdSense and DFP or DFA, the results show generally that it is likely the Google-DoubleClick entity would be able to increase the price of DFP by five or ten percent, even when the amount switching between platforms is very low. That is, Google-DoubleClick would need only a few of the customers it loses after increasing price at DFP to switch to AdSense, instead of switching to a competing platform or not buying at all (given by the critical diversion ratio). It needs even fewer of its total publishing customers to switch to AdSense (given by the critical switching level).[47]

While all the numbers are low due to the nature of the proposed merger, the critical diversion ratios and switching levels implied by the one-sided results turn out to significantly overestimate the two-sided calculations. In other words, the one-sided calculation understates the potential for unilateral price effects from the merger. Whereas the one-sided numbers allow for a small possibility that the price increase would not be profitable, the two-sided calculations suggest that Google is likely to raise prices on DFP even if virtually no customers switched to AdSense.

This creates a "region of error" between the critical diversion ratios. If the actual diversion ratio is below the one-sided figure and above the two-sided figure, it would result in different conclusions depending on the calculation. The one-sided calculation would suggest profitable price increases, the two-sided calculation would not. The same is true for actual switching levels.

The finding of profitable price increases with an effectively zero cross-price elasticity in the two-sided calculation might be surprising. Merging firms in a one-sided world only raise price above its formerly profit-maximizing level if doing so directly causes customers to switch to its other products, i.e. a large enough cross-price elasticity. The reason why zero switching levels are plausible in the two-sided case is because of the indirect network externalities. By raising the price

[47] We present this analysis to demonstrate the effect of two-sided considerations on standard diversion ratio analysis. Whether and under what circumstances it is appropriate to conclude from this sort of analysis that a merger should be blocked is much more complicated issue which we do not address in this article.

TABLE 3 Sensitivity Analysis Across Different Data Assumptions

DFP Assumptions		AdSense Assumptions		5% Critical Price Increase		10% Critical Price Increase	
Price	Margin	Price	Margin	One-Sided	Two-Sided	One-Sided	Two-Sided
7	75%	40	75%	1.17%	0.00%	2.33%	0.74%
7	50%	40	75%	1.17%	0.00%	2.33%	0.74%
7	25%	40	75%	1.17%	0.00%	2.33%	0.74%
5	75%	40	75%	0.83%	0.00%	1.67%	0.01%
5	50%	40	75%	0.83%	0.00%	1.67%	0.01%
5	25%	40	75%	0.83%	0.00%	1.67%	0.01%
3	75%	40	75%	0.50%	0.00%	1.00%	0.00%
3	50%	40	75%	0.50%	0.00%	1.00%	0.00%
3	25%	40	75%	0.50%	0.00%	1.00%	0.00%
5	75%	60	90%	0.46%	0.00%	0.93%	0.00%
5	75%	60	75%	0.56%	0.00%	1.11%	0.01%
5	75%	60	50%	0.83%	0.00%	1.67%	1.02%
5	75%	40	90%	0.69%	0.00%	1.39%	0.00%
5	75%	40	75%	0.83%	0.00%	1.67%	0.01%
5	75%	40	50%	1.25%	0.00%	2.50%	1.53%
5	75%	20	90%	1.39%	0.00%	2.78%	0.00%
5	75%	20	75%	1.67%	0.00%	3.33%	0.03%
5	75%	20	50%	2.50%	0.00%	5.00%	3.06%
5	50%	40	90%	0.69%	0.00%	1.39%	0.00%
5	50%	40	50%	1.25%	0.00%	2.50%	1.53%

of DFP, DoubleClick loses publishers and, via the feedback effects, also loses advertisers, then more publishers, more advertisers and so on. As the DoubleClick platform shrinks and its value deteriorates, advertisers and publishers switch to other platforms, including Google's AdSense.[48] The increase in profits to Google from this switching, which is driven by changes in platform size rather than changes in relative prices directly, makes the price increase on DFP profitable. Hence,

[48] Customers who are sensitive to platform size, even if perfectly price inelastic, make this switch.

Google benefits from contracting and devaluing the low (absolute) margin DoubleClick platform to increase the relative value of its high (absolute) margin AdSense platform. [49]

Table 3 performs robustness checks and reports critical diversion ratios over a wide range of data assumptions.[50] The same patterns hold for critical switching levels. The two-sided calculation is often zero, and always zero for 5% price increases, given the high markups involved.

In summary, we have demonstrated that one-sided and two-sided calculations yield different critical diversion ratios (and critical switching levels) and open up a region of error between them. The size of this region of error depends on the data and externality assumptions, but is non-trivial as a general rule. If the actual diversion ratio falls in between the one- and two-sided calculations, the analyst would give the wrong recommendation about the unilateral price effects of the proposed Google-DoubleClick merger.

VI. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Traditional methods for evaluating mergers between single-sided firms can lead to erroneous conclusions when mergers involve MSPs. The direction of the bias depends on the analytical technique used, with traditional demand estimation tending to draw markets too narrowly and overstating competitive effects, and Lerner Index based calibration methods drawing markets too broadly and understating competitive effects. The direction of bias can also vary with the degree of platform asymmetry and the particular set of products that are subject to price increases. When mergers involve MSPs, the correct analysis must account for the indirect network effects between the multiple sides and the consequent effect on prices and output for the multiple sides. Failing to do so can lead to material mistakes as we have shown through the simulations reported above.

Although we have focused on the numerical mistakes that can result from the application of one-sided methods to markets with multi-sided businesses, it is important to emphasize that most profound analytical mistake in this context results from settling on a one-sided market when the products involve are fundamentally two-sided. This has been the case in, for example, the interchange fee cases in which competition authorities have focused on the impact of the fees paid by merchant acquirers to card issuers in a merchant acquiring market, thereby ignoring the fact that the merchant and cardholder sides are inextricably interlinked.[51]

---

[49] The fact that Google can profit from DFP price increases that are purely contractionary suggests that other non-price methods of shrinking the relative size of the DoubleClick platform can also be profitable. Quality degradation is an example. More likely, however, Google could tie intermediation services to its tools software on the DoubleClick platform, as it currently does on AdSense. This would foreclose third party networks and direct sales channels for intermediation, and effectively convert DoubleClick from a low-margin platform to a high-margin platform like AdSense.

[50] Except as noted, all other data assumptions remain the same.

[51] See Klein et al. (2006), Evans and Schmalensee (2005a, b).

We are dubious that the light generated by market definition analysis in markets involving MSPs is worth the candle. The proper analysis is difficult and highly dependent on proper modeling and data. Mistakes that result in narrow markets that assume away clear multi-sided effects are quite costly. Although the analysis of unilateral effects is also challenging, our advice to practitioners is to avoid formalistic market definition exercises and to use one's candles to illuminate the effects of the merger on prices and other dimensions of competition.

APPENDIX A: N-PLATFORM AND ASYMMETRIC PLATFORM CASES

We extend the results in the text to the case of asymmetric platforms and to the case of more than two platforms. The Critical Loss equation for two asymmetric platforms is given by:

$$\sum_{s=A,B}\sum_{i=1,2}\left[R_i^s(X_i^s + M_i^s)\left(\frac{\Delta Q_i^s}{Q_i^s}\right) + R_i^s X_i^s\right] = 0 \tag{A.1}$$

The N platform formula differs only in the number of summed terms. This formulation allows different price changes at each platform on each side for use in a variety of questions.

For Actual Loss, consider again the isoelastic demand system given by

$$q_i^s = \mu_i^s + \alpha_i^s q_i^r - \delta_i^s q_r^j - \beta_i^s p_i^s + \gamma_i^s p_j^s \tag{A.2}$$

for i,j = 1, 2, i ≠ j and s = A, B, where lower case q's and p's represent log values. Assume all parameters are positive (though this need not be the case) and $\alpha_i^s > \delta_j^s$, $\beta_i^s > \gamma_j^s$, $\alpha_i^s - \delta_j^s < 1$ and $\alpha_i^s + \delta_j^s < 1$ to prevent the two-sided effects from exploding.

Since this a demand system, ownership patterns do not matter for estimation. Also, if there are two platforms that both sell to side A, it is not necessary that they both sell to side B. The two "side B's" can refer to completely different goods. It may also be that one platform sells to sides A and B, whereas another firm sells only to side A, and a third firm only sells to side B.

The reduced form value of q_s^j as a function only of parameters and prices, accounting for all the feedbacks in the system, is:

$$q_i^s = [\Gamma_i^s(\alpha_j^s\theta_j^r - \delta_j^s\theta_i^r + \theta_j^s) + \Psi_i^s(\alpha_i^s\theta_i^r - \delta_i^s\theta_j^r + \theta_i^s)] / \Omega_i^s \tag{A.3}$$

where

$$\begin{aligned}
\theta_i^s &= \mu_i^s - \beta_i^s p_i^s + \gamma_i^s p_j^s \\
\Gamma_i^s &= -\alpha_i^s\delta_i^r - \delta_i^s\alpha_r^j \\
\Psi_i^s &= (1 - \alpha_j^s\alpha_i^r - \delta_j^s\delta_i^r) \\
\Omega_i^s &= \Psi_i^s(1 - \alpha_i^s\alpha_i^r - \delta_i^s\delta_j^r) + \Gamma_i^s(\alpha_j^s\alpha_i^r + \delta_j^s\delta_i^r)
\end{aligned}$$

Totally differentiating, and replacing all dp_i^s with X_i^s and dq_i^s with L_i^s , we derive the Actual Loss formula:

$$\begin{aligned} L_i^s = & (\Gamma_i^s \gamma_j^s - \Psi_i^s \beta_i^s) X_i^s / \Omega_i^s \\ & + (-\Gamma_i^s \beta_j^s + \Psi_i^s \gamma_i^s) X_j^s / \Omega_i^s \\ & + [\Gamma_i^s(\alpha_j^s \gamma_j^r + \delta_j^s \beta_i^r) + \Psi_i^s(-\alpha_i^s \beta_i^r - \delta_i^s \gamma_j^r)] X_i^r / \Omega_i^s \\ & + [\Gamma_i^s(-\alpha_j^s \beta_j^r - \delta_j^s \gamma_i^r) + \Psi_i^s(\alpha_i^s \gamma_i^r + \delta_i^s \beta_j^r)] X_j^r / \Omega_i^s \end{aligned} \tag{A.4}$$

The analyst then compares Actual Loss and Critical Loss as described in the text. The analysis can further be extended to N platforms. The Actual Loss formula is derived from the following system of 2N equations:

$$q_i^s = \sum_{k=1}^{N} \delta_{ik}^s q_k^r + \theta_i^s \tag{A.5}$$

for i = 1..N; s, r = A,B; s ≠ r, where

$$\theta_i^s = \mu_i^s + \sum_{k=1}^{N} \gamma_{ik}^s p_k^r$$

Quantity at platform i on side s depends on the quantity at each platform *k* (including *i*) on the other side *r* (the δ_{ik}'s), and the price at each platform (including *i*) on side *s* (the γ_{ik}'ss).[52] Substituting each q_i^r equation into each q_i^s equation, the system can be rewritten in matrix form: $\Delta^s \vartheta^s = \Phi^s$ for each side s, where Δ^s is a N×N matrix with *ij*th element λ_{ij}^s equal to

$$\lambda_{ij}^s = 1(i = j) - \sum_{k=1}^{N} \delta_{ik}^s \delta_k^r \tag{A.6}$$

where $1(i = j)$ is an indicator function equal to one when $i = j$ and zero otherwise. Also, ϑ^s is a N-vector with ith element equal to q_i^s, and Φ^s is an N-vector with ith element equal to ϕ_i^s , where:

$$\phi_i^s = \theta_i^s + \sum_{k=1}^{N} \delta_{ik}^s \delta_k^r \tag{A.7}$$

Actual Loss is calculated for each q_i^s by total differentiation, i.e. $L^s = (\Delta^s)^{-1} d\Phi^s$, for s = A,B, replacing each dp_i^s with X_i^s in the $d\phi_i^s$ terms.

[52] We have replaced the notation α_i from the two-platform version with δ_{ii}'s and the coefficient on own price from β_i with γ_{ii}s.

To close the model, we note that the first order conditions defining markups in the two-platform symmetric case are the same as those in the 2-platform and N-platform asymmetric cases, since each set of equations is platform specific.

As before, the calculated Actual Loss amounts X_s^j are substituted into the Critical Loss equation and the ACR is checked. (Alternatively, one can fix all parameters except one, and use the ACR = 1 identity to back out the "critical" value of the remaining parameter.) The comparative statics from the symmetric and two-platform cases carry through to the N firm asymmetric case.

Part Three

WEB BASED PLATFORMS

CHAPTER 7

Economics of the Online Advertising Industry

ABSTRACT

Online advertising has grown rapidly and accounts for about 7 % of US advertising spending. It is projected to increase sharply as more consumers spend time online on their personal computers and as additional devices such as mobile phones and televisions are connected to the web. This chapter describes how the online advertising industry works. The industry is populated by a number of multi-sided platforms that singly or in combination facilitate connecting advertisers to viewers. Search-based advertising platforms, the most well-developed of these, has several interesting economic features that result from the combination of keyword bidding by advertisers and single-homing.

I. INTRODUCTION

Online advertising began in 1994 when HotWire sold the first banner ads to several advertisers.[1] Revenue in the United States grew to an estimated $7.1 billion in 2001 or about 3.1 percent of overall advertising spending. The dot-com bust destroyed or weakened many of the early online advertising industry players and reduced the demand for online advertising and related services.

The industry regained momentum by 2004 as the business model for "Web 2.0" came together.[2] A number of businesses such as Advertising.com, Google and ValueClick emerged that facilitated the buying and selling of advertising space on web pages. Many web sites settled on the traditional "free-tv" model: generate traffic by giving away the content and sell that traffic to advertisers. Of the 20 top web sites, based on unique visitors, in the United States, 14 generate the preponderance of their revenues from the sale of advertising inventory—the eyeballs that view

---

[1] Barbara K. Kaye and Norman J. Medoff, *Just A Click Away: Advertising on the Internet* (Massachussetts: Allyn and Bacon, 2001).

[2] "Hundreds of Internet companies have emerged since the dot-com crash, looking to capitalize on a resurgent online advertising market. Companies in this new wave -- known as Web 2.0 -- have focused on online collaboration and sharing among users. They hope to attract millions of users and become the next YouTube, which was acquired by Google Inc. earlier this year for $1.65 billion." See *Is 'Web 2.0' Another Bubble?* The Wall Street Journal, December 27, 2006.

space allocated for promotions—to advertisers.[3] In 2007, advertisers in the United States spent more than $21 billion advertising on websites, about 7 percent of all U.S. advertising spending.[4]

The portion of advertising that is done online is expected to increase significantly over time as web-based content and services expand and people are able to access the web through more devices such as mobile telephones and televisions. The valuations that the capital markets are placing on businesses related to online advertising are consistent with this prediction. Google has had a nearly five-fold increase in its market value from August 2004 when it was valued at $29 billion to $136 billion in April 2008. During 2007 several companies in the online advertising market were purchased at multiples of 10-15 times annual revenues.[5]

The online advertising industry burst into the public eye in 2007. Google's sky-rocketing stock price and its forays into industries such as word processing software, online payments, and mobile telephones drew significant attention. More than 500 articles on Google appeared in the *New York Times, Wall St. Journal* and the *Financial Times* during the year. The U.S. Federal Trade Commission and the European Commission launched in-depth antitrust investigations into Google's acquisition of DoubleClick, which provides software technology and services to online advertisers and publishers.[6] Privacy concerns also came to the fore in 2007 as consumers, government agencies and the media started focusing on the massive amount of personal data that online advertising companies were storing and using.[7]

This chapter describes how the online advertising industry works, focusing on several economic aspects of this business. Although the online advertising industry has revolutionized many aspects of an age-old business, it is important to understand, as we present in Section 2, that the new industry has much in common with the old. The unique features of online advertising include the use of Internet-based technologies and data collection mechanisms to target and track specific individuals

---

[3] Of these 14, five also use a subscription model to supplement revenues. Out of the remaining six, four use the merchant model, one uses the auction model (eBay.com), and one is a not-for-profit (wikipedia.org).

[4] eMarketer, *US Online Advertising: Resilient in a Rough Economy,* March 2008. Estimates of advertising spending, especially online advertising spending, should be viewed as indicative.

[5] Google announced in May 2007 that it would purchase DoubleClick for $ 3.1 billion which is more than 10 times DoubleClick's revenues according to one account. Louis Story and Miguel Helft, *Google Buys an Online Ad Firm for $3.1 Billion,* New York Times, April 14, 2007. Microsoft purchased aQuantive at a multiple of about 13. Peter Galli, *Microsoft's aQuantive Buy Shows Big Ad Plans,,* eWeek.com, May 18, 2007. Yahoo paid some $680 million for 80 percent share of Right Media which generated about $35 million in revenues in 2006. Michael Liedtke, *Yahoo snaps up Right Media for $680M* , USA Today, April 30, 2007.

[6] European Commission Press Release, *Mergers: Commission opens in-depth investigation into Google's proposed take over of DoubleClick,* November 13, 2007; Google SEC Filing, *Form 8-K,* sec.gov, May 29, 2007. Also, the United States Senate held hearings on this acquisition. See *An Examination of the Google-DoubleClick Merger and the Online Advertising Industry: What Are the Risks for Competition and Privacy?,"* Senate Judiciary Committee, September 27, 2007. The FTC cleared the transaction on December 20, 2007. *See Statement of Federal Trade Commission concerning Google/DoubleClick*, FTC File No. 071-0170, http://www.ftc.gov/os/caselist/0710170/071220statement.pdf.

[7] Vidya Ram, EU Turns Spotlight On Google, Forbes, May 28, 2007; Steve Lohr, *Google Deal Said to Bring U.S. Scrutiny,* New York Times, May 29, 2007; Darren Waters, Google privacy policy 'is vague', BBC News, May 31, 2007.

and to automate the buying and selling of advertising inventory. Like modern finance, online advertising relies heavily on advanced economic and statistical methods. These topics are discussed in Section 3, which focuses on search-based advertising- the most well developed part of online advertising business to date- and Section 4, which examines online display advertising, a rapidly evolving part of the business that involves advertising on any web sites that draw traffic. The online advertising industry is highly complex, undergoing a series of rapid changes, and could well result in a high degree of concentration, if not monopoly, in the intermediation of advertising inventory and the control of personal data. Section 5, presents concluding remarks, and explains why the online advertising industry will remain at the center of public policy debate for many years to come.

II. THE ADVERTISING BUSINESS

Advertising is designed to promote the sale of a product or service. It has been around in some form since ancient times and occurs in many cultures. The business of presenting advertisements to people became enormous during the 20th century with the development of various methods of mass communication and the perfection of the advertiser-supported model for delivering content. Advertising spending worldwide is over $625 billion a year, a number that exceeds worldwide spending on wireless voice communication.[8]

A. Brief History

Outdoor advertisements were some of the earliest methods of promoting sales, with signs appearing in Babylonia as far back as 3000 BC. An ad found in the ruins of Pompeii points travelers to a tavern in a nearby town. A key innovation in the history of advertising was the insertion of ads into media that attracted viewers. The first newspaper ad was reported to have occurred in 1672, offering a reward for the return of 12 stolen horses. The *Boston News-Letter* began carrying ads in 1704.[9]

The advertising industry has developed at least in part as a result of media companies realizing—as web sites have recently—that a profitable business model involves using content to attract viewers and selling access to those viewers to advertisers. The U.S. magazine industry settled on this "two-sided" model in the late 19th century.[10] One of the leading publishers dropped its maga-

[8] Stuart Elliott, *Forecasters Say Madison Avenue Will Escape a Recession, Just Barely,* The New York Times, December 4, 2007. Stephen Minton, *Worldwide Telecom Spending 2007–2011 Forecast: Worldwide Telecom Black Book, 2007,* IDC, November 2007.

[9] *Advertising,* Microsoft Encarta Online Encyclopedia 2007, http://encarta.msn.com/encyclopedia_761564279/Advertising.html.

[10] For an introduction to two-sided business models see, David S. Evans and Richard Schmalensee, *Catalyst Code: The Strategies of the World's Most Dynamic Companies* (Massachusetts: Harvard Business School Press, 2007). The seminal economics paper in this area is Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. OF EUR. ECON. ASS'N 990 (2003).

zine price sharply to increase circulation, and instead earned revenue from selling advertisements. Revenue and profits increased from this pricing innovation. Most magazine publishers quickly followed, and today that is how most earn their profits. The radio industry initially struggled with a subscription-based model, but several stations discovered the power of advertising and the rest quickly followed. Television followed the same path.[11]

Advertising agencies emerged in the mid 19th century as brokers between newspapers and advertisers. The first agency in the United States started in 1841. Its agents bought large amounts of newspaper advertising space at a discounted price and then resold it to advertisers. At first the advertisers designed the ads and the agency just placed them, but later on advertising agencies started designing the ads and providing other services. The business model that eventually developed involved giving creative and marketing services away in return for commission on the media buying.

B. Role of Advertising

Although all advertising is ultimately designed to generate the sales of goods and services, it does so in different ways. Some advertising is designed to generate sales directly by identifying "leads". People who look up a type of service in the yellow pages, for example, are generally interested in purchasing that service. The paid listings and advertisements in the yellow pages are designed to encourage solid sales prospects to patronize the advertiser's business. Other advertising is informative. It provides consumers with information about prices and products, which they can use to make purchasing decisions. Newspaper ads for supermarkets that list sale items and their prices are an example. Still other advertising is about branding or altering people's perceptions about a product or service. The "Visa Is Life" television advertisements are an example of this. The lines between lead-generation, information provision, and branding are blurry. The distinctions are important for the discussion below, however, because online advertising has provided especially innovative technology for generating leads.[12]

C. Pricing and Business Models

The most common pricing method in the advertising industry is based on cost per viewer, often expressed in terms of the cost of reaching a thousand viewers (CPM). Newspaper, radio, and television advertisements are typically sold based on estimates of the number of people with certain demographic characteristics who will view an ad that has been placed in one of those media outlets. Television ad rates in the United States, for example, are largely determined by

[11] Evans and Schmalensee, *supra* note 10; Microsoft Encarta supra note 9

[12] For a general introduction to the economics of advertising see Kyle Bagwell, *The Economic Analysis of Advertising,* Mark Armstrong and Rob Porter (eds.), Handbook of Industrial Organization, Vol. 3, North-Holland: Amsterdam, 2007, pp. 1701-1844.

Nielsen Media Research's data on demographics and what is being viewed. Broadcasters and ad agencies negotiate prices based on Nielsen's numbers, and the outcomes of these negotiations in turn determines which programs remain on the air.[13] Furthermore, contracts between advertisers and TV networks usually include a rating guarantee. Should ratings of the program in which the ad is shown fall short of the agreed level, TV networks would provide extra ad time to the advertisers.[14]

Traditional media that use content to attract viewers have adopted two different models. In the subscription/advertising model the publisher charges viewers a fee to obtain access to the content, and advertisers a fee to obtain access to the viewers. Many newspapers and magazines follow this model. They then balance the demand from advertisers and subscribers to maximize revenues. Some magazines, e.g., the *Economist*, have adopted reader-friendly strategies with high reader fees but sparse advertising. Others have adopted advertiser-friendly strategies, e.g., *Vogue*, with lower reader fees and more advertising, some of which makes reading the magazine difficult. In the free-media model the publishers do not charge viewers for access to the media at all, and in fact try to distribute the media as widely as possible. They earn all of their revenues and profits from the sale of advertisements. Free radio and television have embraced this model in the United States. However, there are many free newspapers and magazines that have adopted the free-TV model. Pay-television and satellite radio have adopted the mixed subscription/advertising model. These different business models are now better understood as a result of the work on multi-sided platforms.[15]

D. The Online Advertising Industry

The online advertising industry concerns buying and selling advertising space that is accessed by viewers through the Internet. Industry observers often divide the on-line advertising industry into: (1) "search advertising" that appears on search-results pages; (2) "display advertising" that appears on non-search web pages; (3) classified listings that appear on web sites; and (4) Internet e-mail

[13] Franklin M. Fisher, John J. McGowan, and David S. Evans, *The Audience Revenue Relationship for Local Broadcast Stations,* Bell Journal of Economics (Autumn 1980); *Nielsen Starts Watching the Ad Watchers,* Journalism.org, November 2, 2006; *Collecting and Processing the Data,* Nielsen Media Research, http://www.nielsenmedia.com.

[14] For example, "[I]f NBC Universal [did] not deliver the viewers it [had] promised advertisers, it would have to offer them compensatory commercial time…" (Stuart Elliott, *Olympic-Size TV Audience for the Athens Games?,* New York Times, August 13, 2004. Also "ABC, CBS, NBC, Fox and the CW network sold $9.3 billion in prime-time ads for this season. In the process, they sold about 80 percent of their time, holding back some to give advertisers should ratings fall short of guarantees." (Meg James and Alana Semuels, *Lower ratings could pinch TV ads,* Los Angeles Times, December 12, 2007.

[15] Rochet and Tirole, *supra* note 10; Simon P. Anderson & Jean J.Gabszewicz, *The Media and Advertising: A Tale of Two-Sided Markets,* Handbook of Cultural Economics, Elsevier, 2006; Simon Anderson and Stephen Coate, Market Provision of Broadcasting: *A Welfare Analysis*, Review of Economic Studies, Vol 72, No. 4, October 2005, pp. 947-972.

TABLE 1 US Ad Spending, 2002 - 2006 (Billions). Display includes display ads, rich media, and sponsorship. Rich media includes interstitials.

	Search	Display	Classifieds	E-mail	Other	Total
2002	$0.90	$3.42	$0.90	$0.24	$0.54	$6.0
2003	$2.56	$2.99	$1.24	$0.22	$0.29	$7.3
2004	$3.74	$3.65	$1.73	$0.19	$0.29	$9.6
2005	$5.13	$4.25	$2.13	$0.25	$0.75	$12.5
2006	$6.76	$5.41	$3.04	$0.34	$1.35	$16.9
% Change: 2002-2006	651%	58%	238%	41%	150%	182%

Source: IAB Internet Advertising Report, 2002-2006 Full-Year results. Available at http://www.iab.net/insights_research/1357.

based advertisements. Table 1 reports U.S. advertising spending for 2006 in these categories and growth between 2002 and 2006.

Search-based advertising accounted for the largest portion with 40 percent followed by display-related advertising with 32 percent in 2006 (of which 22 percent was display advertising, with rich media and sponsorship accounting for the remaining 10 percent.) All segments have grown significantly in the last few years with search having the highest growth rate.

In many ways, on-line advertising is similar to traditional advertising. Publishers use content to attract viewers and then sell advertisers access to those viewers. Advertisers can display text (like classifieds), graphics (like magazines) and video (like television) ads in the space supplied by the publishers. On one level, one can think of the web as just adding more advertising inventory, much like displaying ads on televisions in the back of taxis. Indeed, in some ways the introduction of online advertising was a less radical innovation than the introduction of other media. After all, television enabled advertisers to reach mass audiences with video ads while the web is relying on quite traditional methods of presentation.

Three radical innovations, however, distinguish online from off-line advertising. The first has transformed the service obtained by the advertiser: the Internet provides a highly efficient mechanism for delivering ads to individual users and collecting information for targeting ads to those users. The second has transformed the process of buying and selling advertising space: the Internet has enabled the development of more efficient intermediation markets for advertising—the keyword bidding system used for search and contextual advertising is the most mature example. The third is leading to economies of specialization: traditional publishers provide content for attracting viewers and sell advertising space to advertisers; online publishers are increasingly turning the selling of advertising space over to specialized advertising platforms such as Google or advertising.com.

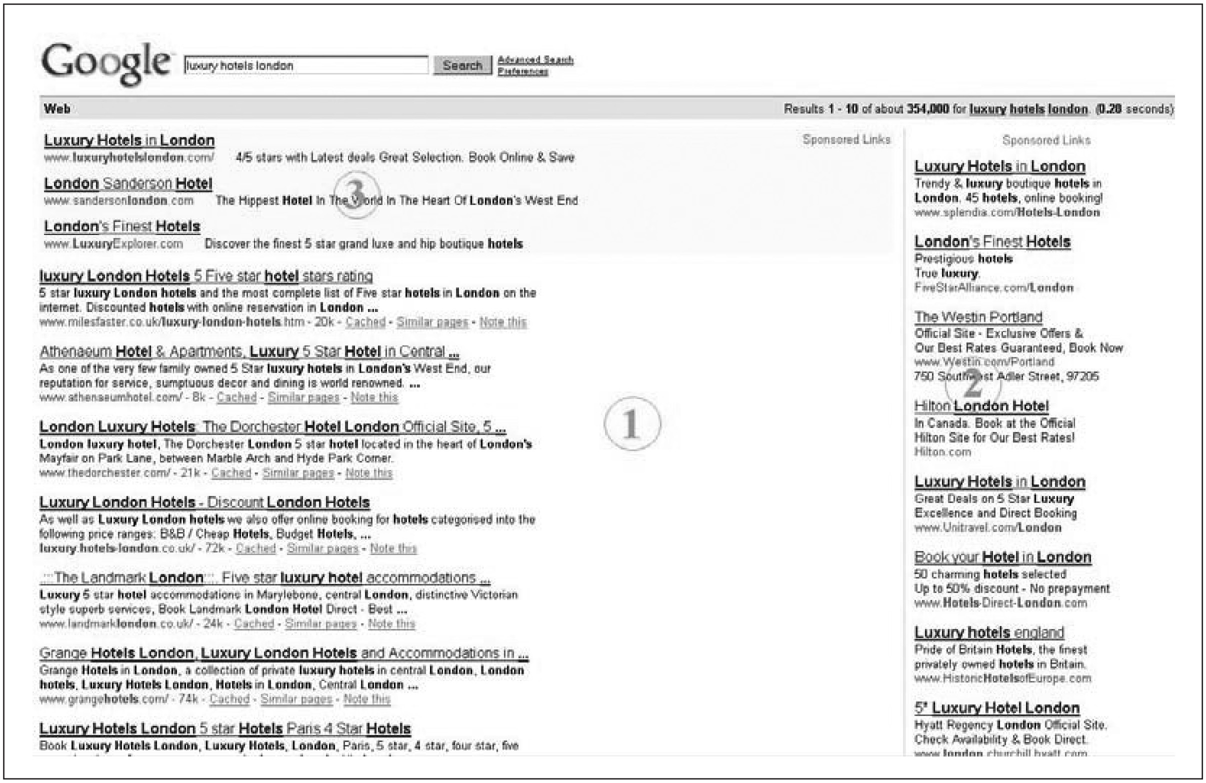
As more advertising moves to Internet-connected devices, these innovations will dramatically alter the advertising ecosystem. These innovations are mainly affecting search and display advertising, which are the focus of this chapter.

III. ADVERTISING ON SEARCH RESULTS PAGES

When you enter a search query using one of the commercial search engines you will often see the web page divided into up to three areas as shown in Figure 1.

The left-hand side of the screen displays the "organic search results". These are based on an index of the world-wide-web maintained by the search engine provider and selected based on algorithms that rank their likely relevance to the search query term(s). The right-hand side of the screen displays "paid search results," which are listings sold by the search engine provider to advertisers. The top-left-hand side of the screen above the organic search results may also include paid search

FIGURE 1 A search on Google for "luxury hotels London" shows the resulting web-page divided into three areas: 1) organic search results, 2) paid search results, and 3) more paid search results.



results for some search engines. A search query may generate a series of pages of search results, and each page may have ads on the right-hand side if there have been buyers for the space.

A “search-based advertising platform” (search-ad platform for brevity) attracts viewers to its pages largely by displaying the organic search results from its search engine. It allocates a portion of the page for the purpose of selling advertising space, divides this place into slots (there are typically 8-10 per page) which it sells to advertisers. There are two key technological innovations that underlie this process, both of which depend on advanced economic and mathematical methods, and which ultimately help determine the nature of the market structure.

A. Technological Underpinnings of Search-Based Advertising

Search-ad platforms use a “keyword bidding system.” Advertisers bid on search query terms known as keywords. They can bid on individual keywords as well as combinations such as “hotel”, “hotel Boston”, “luxury hotel Boston”, and “hotel Beacon Hill”.[16] The major commercial search-ad platforms use a second-price auction with a reserve price for this auction.[17] The price is based on the charge for each time an Internet user clicks on the ad (“cost-per-click” or CPC). All else equal, a higher bid price will secure a higher slot (one closer to the first slot at the top of the first page).

The bid itself does not, however, necessarily determine the slot that an ad is placed in, which brings us to the second technological innovation. The search-ad platforms want to maximize the revenue they receive from selling slots. Since they have chosen to charge based on CPC they need to take into account the number of clicks that an ad will receive. They may earn more profits by putting ads with lower CPC bids in higher slots if doing so generates more clicks than ads with higher CPC bids. To maximize revenue the search-ad platform therefore needs to estimate the “click-through-rate” (CTR) for a search ad bid and allocate the slots to bidders to maximize revenue. In addition, the platform must consider the ultimate relevance of ads to viewers. An ad that is good at attracting clicks but does not result in sales to users will likely leave users dissatisfied with seeing less relevant ads and advertisers dissatisfied with having paid for ads that did not lead to sales. Google does this by estimating a “quality score” for each bid that reflects the expected CTR and the relevance of the

[16] When users bid on keywords, those keywords are set to a certain match option. These options include a broad match (which displays the ad when customers search for words in the keyword list in any order and possibly with other terms), a phrase match (which displays the ad when a customer’s search query includes all keywords in the exact order given, even if the query has other terms that precede or follow the phrase), and an exact match (which makes an ad eligible when a search includes the specific keyword or phrase, in order, and without any other terms in the query). Negative keywords can also be added so that an ad will not be displayed if a search query contains a negative keyword.

[17] See Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz, Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords, American Economic Review, v. 97(1), March 2007, pp.242-259.; Microsoft, What’s the Cost?, http://advertising.microsoft.com/search-advertising/cost. Also see Hal R. Varian, Position Auctions, Working paper, University of California, Berkeley, 2006.

ad to a searcher. Estimating these types of quality measures is especially difficult for advertisers and keyword combinations for which the search-ad platform has no experience.[18]

The keyword bidding process and the quality score algorithm together determine both the CPC advertisers pay and the slots they receive. Search-ad platforms sometimes provide bidders guidance on what they would have to pay to get particular slots. For example, Google estimates that bidders would have to pay an estimated $3.04 to get the third slot for "luxury hotels London" and $0.05 to get the third slot for "competition economists."[19]

These technologies affect the market structure in two ways as we will see below. First, the keyword bidding system can give rise to demand-side scale economies. There is essentially a liquidity effect arising from larger platforms having thicker markets for keywords. Second, platforms that have superior technologies can earn more from additional searchers and therefore bid more for traffic thereby accelerating positive feedback effects. Before discussing these features in more detail we survey the current state of competition in the search-based advertising business.

B. Market Structure of Search-Based Advertising

Because search-based advertising is two-sided, one needs to examine the position of search-ad platforms on their ability to generate search traffic as well as their ability to sell that search traffic to advertisers. Considering these two dimensions, we begin by looking at the current structure of the business and its evolution over time.

1. Market Structure in 2007

In the United States the search-based advertising business has three major players, as well as some fringe firms. The shares of search traffic are reported in Table 2. There is a consensus in the industry that the larger platforms realize higher revenue per search than small platforms.[20] Consequently, the shares based on revenue are more highly skewed than the shares based on search traffic.

[18] The exact methodology of each search ad platform and the precise dates of implementation are not public but there have been significant differences across platforms. Google appears to have had a quality score since at least August 2005 (http://adwords.blogspot.com/2005/12/new-addition-to-quality-score.html). Yahoo does not appear to have relied on these quality measures until its Panama platform was implemented in February 2007. Microsoft used Yahoo's platform until it launched its own search ad platform in the United States in May 2006. Microsoft's platform appears to have always had some form of quality scoring. See http://blogs.msdn.com/adcenter/archive/2006/08/23/715430.aspx discussing an "enhancement" to "existing guidelines around relevance and quality" in April 2007. There may of course still be significant differences in the effectiveness of the quality scores across platforms.

[19] See Varian (2006) for an analysis of the equilibrium bidding for slots. Also see Michael Schwarz and Konstantin Sonin, *Efficient Actions in a Dynamic Auction Environment* (2005).

[20] We discuss evidence on this below.

TABLE 2 US search engine share by search queries.

Platform	Share of Search Traffic, Dec 2004	Share of Search Traffic, Dec 2005	Share of Search Traffic, Dec 2006	Share of Search Traffic, Dec 2007
Google	35.6%	40.5%	50.8%	61.3%
Yahoo	32.7%	29.9%	26.6%	20.2%
MSN	16.7%	14.5%	9.4%	8.2%
Time Warner / AOL	9.6%	8.7%	9.1%	6.7%
Ask	5.4%	6.4%	4.2%	3.6%
HHI	***2,735***	***2,864***	***3,473***	***4,291***

Source: comScore MediaMetrix.
Dec. 2004 – Dec. 2005: Custom query prepared by comScore, January 2007.
Dec. 2006 – Dec. 2007: qSearch 2.0 Key Measures Report, December 2006 and December 2007.
Notes: Shares are calculated from searches conducted on the above five web properties only. The share for the Time Warner Network may be slightly overstated as searches conducted within its websites, such as Mapquest, are included. The HHI estimates are slightly overstated as only the shares of the top 5 firms are considered, but the trend in the estimated HHIs should be indicative of that of the actual HHIs.

The industry is highly concentrated with an HHI of over 3,000 based on search traffic and higher based on advertising revenue.[21] The size distribution of firms is also highly skewed. Measured by 2006 advertising revenue, the largest platform (Google) is nearly three times as large as the second-largest platform (Yahoo).[22] The rankings by search traffic are slightly less skewed.

Most consumers have, or can easily obtain a search engine. The typical computer comes with a search toolbar preinstalled in the browser. A 2007 study examined PCs sold by OEMs who accounted for at least 1 percent of U.S. home and small office sales in 2006. All of the PCs had a search toolbar preinstalled, with Google having the greatest number of installations.[23] It is also easy

[21] The HHI is a standard measure of concentration used in merger analysis. It ranges from a high of 10,000 for a pure monopoly to 0 for a perfectly competitive industry. Industries with HHIs in excess of 1,800 are considered to be sufficiently concentrated that US antitrust authorities look closely at mergers in these industries. See Federal Trade Commission, *1992 Horizontal Merger Guidelines[with April 8, 1997, Revisions to Section 4 on Efficiencies*], http://www.ftc.gov/bc/docs/horizmer.shtm.

[22] Neither Google nor Yahoo break out their search advertising revenues from their contextual advertising revenues, so this ratio is for both search and contextual advertising combined. Google's worldwide 2006 search and contextual advertising revenue is estimated at $10.5 billion and Yahoo's worldwide 2006 search and contextual advertising revenue is estimated at $3.7 billion. Estimates come from the Lehman Brothers Internet Data Book, April 2007, pp. 81, 85. The Google to Yahoo ratio based on these revenue figures is about 2.9, while the ratio for search traffic of Google to Yahoo's in November 2007 was 1.7.

[23] Google was found on three of the seven PCs, Yahoo! on two, AOL on one, and Windows Live on one.

to add additional search toolbars, such as the Windows Live Toolbar and the Yahoo! Toolbar. Thus, consumers can use multiple search engines if they want and it is relatively easy to do so.[24] Looking among users of Google, Yahoo and Microsoft's search engines over the course of one month, more than a third used more than one search engine. But the extent of multihoming differed significantly across platforms—only about 40 percent of users of Google used one of the two other search engines over the course of a month, while the comparable figures were over 70 percent for Yahoo users and 80 percent for Microsoft users.[25] That could come from their using multiple engines for the same search, or different engines for different searches.[26] Nevertheless, most people I have asked (acquaintances plus several polls of audiences) use a single search engine primarily and tend to use other ones for idiosyncratic reasons. Moreover, only rarely in my experience do people use multiple search engines for a single query. Thus in the remainder of the paper I will assume that users "single-home" on a search-ad platform using the terminology from the two-sided platform literature.

On the advertiser side, multihoming on search ad platforms varies significantly across advertisers. Although representative data are not publicly available, discussions I have had with industry participants indicate that all of the largest advertisers, such as Amazon and WalMart, will advertise on all of the three major search ad platforms. The value of a click from one search engine is independent of the value of a click from another search engine since these are different lead opportunities if people single-home for a given search. All else equal, there is therefore no reason to pay for clicks only from a single search engine.[27] The major incentive not to use an additional search-ad platform is the cost of setting up that platform and monitoring ad campaigns on it. Those costs, which are largely fixed, will be amortized over fewer clicks on a smaller search ad platform. Discussions I have had with industry participants indicate that medium and small search advertisers will use Google uniformly but are less likely to use Yahoo and, especially, Microsoft. In

---

[24] Google has programmed its search toolbar to make it difficult for users to easily switch. When a user attempts to change the default search, Google blocks the switch by default and briefly displays a notification in the bottom right corner of the screen. If the user clicks on the notification before it disappears, a popup window provides the user with the option to disable Google's "default search protection". This does not appear to be the case for Yahoo or Microsoft.

[25] comScore MediaMetrix, Digital Calculator Report, January 2008. Note that the extent of multihoming for users overall is lower than for any individual platform because a given user that multihomes between say Google and Yahoo will be counted as multihoming twice in the individual platform multihoming figures and only once in the overall figure.

[26] There are also "meta" search engines such as Excite, Dogpile, and Metacrawler that deliver results from several engines. These are not widely used as of 2007.

[27] This statement is not meant to suggest that there are no substitution possibilities from the standpoint of the advertiser. Search generates leads which can translate into sales with some probability. However, advertisers have a number of different methods available to them to generate leads of which search-based advertising is only one way. We would expect that advertisers will invest in different methods of lead generation up to the point where the returns are equalized at the margin. They may therefore increase their investment in platforms that have lower CPCs than platforms that have higher CPCs for given keywords.

addition, advertisers both large and small that do use multiple platforms will tend to place fewer bids on the smaller platforms. In aggregate, this leads to significantly more bids on Google than on Yahoo or Microsoft.[28] We discuss the fixed costs of maintaining search advertising campaigns in more detail below because they can have important implications for positive feedback effects.

2. Evolution of Market Structure

Search-based advertising started in 1995 when Infoseek, one of the earlier search engines for the web, began to target banner ads in their system to the keywords users entered. The CPC model was introduced in a deal Proctor and Gamble struck with Yahoo in 1996.[29] In 1998, GoTo.com introduced the first clearly marked ads alongside organic search results and charged advertisers for these ads based on CPC. GoTo.com became Overture which in turn became Yahoo's ad search platform. Other search engines quickly followed this model. Google launched AdWords in 2000, which was similar to GoTo but distinguished the ads from the organic search results more clearly.[30]

Thus far there have been two leaders in search-based advertising. Although Infoseek started the search-ad business in 1995 no real leader emerged until 1999. Yahoo held the top spot from 1999 to 2002. Google achieved a higher share than Yahoo of search traffic in the U.S. in 2003 and has held the lead ever since. Google's page rank algorithm, used to decide which results to present to a searcher, has been described as the "crucial part of Google's inner sanctum, a department called 'search quality' that the company treats like a state secret."[31] Table 3 reports the first, second and third players in the US from 1999 to 2007 (the table also indicates whether the firm used its own search engine (S) and ad platform (A) technologies or whether those were outsourced).[32]

---

[28] *Google AdWords Vs. Microsoft AdCenter,* Brick Marketing, October 27, 2007, http://payperclickjournal.com/google-adwords-vs-microsoft-adcenter/10/27/2007/; Henry Blodget, The Real Reason Yahoo's Revenue Per Search Stinks, Silicon Alley Insider, Oct 2, 2007, http://www.alleyinsider.com/2007/10/the-real-reason.html.

[29] See *Yahoo and Procter & Gamble Develop Interactive Traffic Building Promotion*, Google and Yahoo News and Press Releases, May 8, 1996.

[30] For a discussion of the evolution of search-based advertising see John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture* (Penguin Group 2005), in particular Chapter 3; also see Michael Cho et al, Search-based Online Advertising (Working Paper, 2005) people.ischool.berkeley.edu/~hal/Courses/StratTech07/Tech/Tech05/C-report.doc.

[31] Saul Hansell, Google Keeps Tweaking Its Search Engine, New York Time, June 3, 2007. Google has described PageRank, its system for ranking web pages, as the "heart" of its software. (Google, *Our Search: Google Technology*, http://www.google.com/technology/, downloaded December 2007)

[32] An interesting area of exploration concerns differences in market structure internationally. Google is the dominant search provider globally and in many countries. It has reportedly more than a 90 percent share of search traffic in Argentina, Brazil, and Spain. However, it is not the largest player in China where Baidu has 61 percent of the search traffic or Japan where Yahoo! has 50 percent.

TABLE 3 Leaders in search in the US by traffic.
Search Website (S: Search technology powering search website, A: Ad platform)

Place	1999	2000	2001	2002	2003	2004	2005	2006	200
1st	Yahoo (S: Inktomi A: Own)	Yahoo (S: Google A: Own)	Yahoo (S: Google: A: Own)	Yahoo (S: Google, A: Overture)	Google (S: Own, A: Own)	Google (S: Own, A: Own)	Google (S: Own, A: Own)	Google (S: Own, A: Own)	Google (S: Own, A: Own)
2nd	AltaVista (S: Own A: DoubleClick)	AltaVista (S: Own, A: DoubleClick)	Microsoft (S: Inktomi, A: Overture)	Google (S: Own, A: Own)	Yahoo (S: Google, A: Own)	Yahoo (S: Own, A: Own)	Yahoo (S: Own, A: Own)	Yahoo (S: Own, A: Own)	Yahoo (S: Own, A: Own)
3rd	Excite (S: Own, A: Excite's MatchLogic)	Lycos (S: Fast Search & Transfer A: Own)	Google (S: Own, A: Own)	Microsoft (S: Inktomi, A: Overture)	Microsoft (S: Yahoo's Inktomi, A: Yahoo's Overture)	Microsoft (S: Yahoo's Inktomi, A: Yahoo's Overture)	Microsoft (S: Own, A: Yahoo's Overture)	Microsoft (S: Own, A: Own)	Microsoft (S: Own, A: Own)

Notes: Yahoo acquired Inktomi in 2003; In 1999 and 2000, AltaVisa used DoubleClick for banner ad (not ads matched to user search queries).

Source: Data for years 1999-2003 compiled from the following trade press articles: Sherman Fridman, Yahoo's Popularity Continues Strong-Report, Newsbytes New Network, December 22, 1999; Altavista Climbs in Search Referrals, PR Newswire, July 12, 2000; Yahoo Still on Top in Portal Battle, but MSN Creeping Up, Agence France-Press, May 30, 2001; Google Challenging Yahoo as top Global Internet Search Engine: Study, Agence France-Presse, April 30, 2002; comScore Media Metrix Launches Breakthrough System to track Actual Consumer Search Queries, comScore press release, April 28, 2003; Search Sites Play a Game of Constant Catch-Up, New York Times, January 31, 2005. Data for years 2004-2005 from David Hallerman, Search Marketing: Players and Problems, eMarketer, April 2006. Data for years 2006-2007 compiled from comScore reports.

It is useful to focus on the industry since the dot.com bust which marked the collapse of many of the early startups and the emergence of the advertising-supported Web 2.0 industry. Since that time the search-based advertising industry has become increasingly concentrated as Google has outdistanced its rivals. The gap between Google and rest of the search firms has increased substantially—the ratio of the largest firm (Google) to the next largest (Yahoo) has increased from 1.3 to 3.1 over the same period. Table 4 reports the growth rates in search traffic and advertising revenue from 2001 to 2007. In the US, Google has exceeded the growth rates of the second two largest platforms in search in 7 of the last 10 quarters and in revenue in 9 of the last 10 quarters (through the second quarter of 2007).

Google's high share combined with trends indicating that its lead is expanding raise several interesting economic questions: is search-based advertising subject to winner-take-all competition; how easily can more efficient providers displace the incumbent; and to what extent can differentiation on either platform side permit viable competition among several platforms? We turn to these questions next.

C. Economic Factors Affecting Two-sided Market Structure

In many ways search-ad platforms are subject to the same economic considerations as traditional media platforms. They are two-sided businesses based on using bait to attract eyeballs and selling

TABLE 4 US search engine share by search queries.

	2000	2001	2002	2003	2004	2005	2006	2007E
US Search traffic (billion queries)	n/a	n/a	n/a	n/a	49.0	57.5	68.0	86.4
US paid search revenue ($ millions)	$109	$299	$927	$2,544	$3,850	$5,142	$6,970	$8,288

US Search Traffic comes from comScore. Source for US Paid Search Revenue: David Hallerman, Search Marketing: Counting Dollars and Clicks, eMarketer, April 2007. The 2007 revenue figure is an estimate.

access to those eyeballs to advertisers. However, several features of the technologies underlying search-ad platforms give rise to unique features.

1. Pricing of Keywords

Consumers appear, at least based on anecdotal evidence, to use a single platform for a search query.[33] This is akin to consumers who use a single yellow pages directory for looking up merchants. Advertisers often use multiple search-ad platforms just as they put ads for the same product in magazines that are likely to reach different consumers, such as *Vogue* and *Popular Mechanics*. As a result of these facts, market forces do not necessarily lead to the same CPC for a given keyword across search-ad platforms. Neither consumers nor advertisers are making marginal substitution decisions between a given keyword on different platforms. If the CPC for "flat panel televisions in Chicago" was higher on platform 1 than on platform 2, a Chicago television retailer would still use both platforms so long as the CPC was worth the value of the lead generated.[34]

The CPC is ultimately determined by the keyword bidding auction on each platform. Those auctions could result in similar CPCs for given query terms if there were the same bidders, the auction rules were similar, and the values of leads for different platforms were similar. We would expect, as noted above, that platforms that attract fewer bidders for keywords would tend to have lower CPCs for those keywords. We would also expect that platforms that have less efficient auctions, or generate less valuable leads, would have lower CPCs for given keywords.

[33] There would not seem to be any economic or technological obstacle to consumers using multiple search engines. They are most likely to use the search engine that they expect will provide the best search results which may be some combination of organic results and ads. However, it would seem possible that a market structure could arise with differentiated platforms in which consumers find it useful to rely on several platforms or in which they have greater incentives to use a meta-search platform that combines results from several platforms. Therefore the existence of single-homing may be endogenously determined by the market structure.

[34] In reality there are a variety of ways to generate leads, beyond search-ad platforms, and there may be some substitution between these different methods. Moreover, there may be diminishing returns to the value of leads for a variety of reasons including diseconomies of scale in production and distribution.

In fact, there are significant differences in CPCs for keywords across the search-ad platforms. None of the platforms reports these measures, but industry estimates generally place Google significantly above Yahoo. One estimate places Google's worldwide CPC at around $2.00, nearly three times its estimate of Yahoo's CPC of around $0.75.[35]

2. The Role of Indirect Network Effects

It might appear that indirect network effects are insignificant for search-ad platforms, if they are present at all.[36] An advertiser only pays when a consumer clicks on his ad. That value does not depend on whether any other consumer on the platform clicks on the ad. The advertiser should therefore be indifferent to using a platform with few or many searchers so long as the value exceeds the cost of each click the advertiser gets. It might appear that searchers do not benefit in any obvious way from other searchers. So long as they obtain the information they are looking for, they do not care whether the search provider has many or few searchers. They probably value search-ad platforms that have more ads. But given that advertisers are indifferent, the density of ads on search pages should not vary depending on the number of searches. This view, however, ignores key features of transaction platforms that would appear to lead to strong indirect network effects.

Search-ad platforms are similar to other transaction platforms that seek to match buyers with sellers and consummate trades. With more buyers there is a higher likelihood that a seller will find a suitable match that will lead to a beneficial sale, and with more sellers there is a higher likelihood that a buyer will find a suitable match that will lead to a beneficial purchase. The importance of

[35] See *Not Out of the Woods But A Step In The Right Direction,* Bear Stearns, October 17, 2007. Other estimates of revenue per search (RPS), which is equal to the CPC times the CTR for each search (not the CTR for each advertiser), place Google substantially above Yahoo. Mark Mahaney, an analyst in Citigroup, estimated that in 2006 Google made 4.5 cents to 5 cents on every search, while Yahoo generated only 2.5 cents to 3 cents a search (See Miguel Helft, *A Long-Delayed Ad System Has Yahoo Crossing Its Fingers,* New York Times, February 5, 2007); Caris & Co. analyst Tim Boyd estimates that Yahoo made on average between 10¢ and 11¢ per search in 2006, and Google made between 19¢ and 21¢ per search (See Catherine Holahan, *Why Yahoo's Paname Won't Be Enough,* BusinessWeek, Dec 26, 2006); Justin Post, an analyst with Merrill Lynch, estimated that in 2006, each US search generates 4¢ for Yahoo and 11¢ for Google (See Robert D. Hershey Jr., *Sunny and Gloomy Signs at a Web Crossroads,* New York Times, Nov 19, 2006.); A recent article states that most analysts estimate that Yahoo!'s RPS is about 30 percent less than Google's (See Henry Blodget, *The Real Reason Yahoo's Revenue Per Search Stinks*, Silicon Alley Insider, Oct 2, 2007.). The differences between these estimates is likely due to whether one looks at searches only within the U.S. or worldwide, and whether one looks at only searches for which there are ads or all searches. But within each estimate, Google is placed significantly above Yahoo.

[36] Although search-ad platforms have to make investments in developing search engines and other technologies it does not appear that there are significant scale economies that would, by themselves, limit the market to one or just a few players. Search-ad platforms are readily scalable by adding servers and communication. Eisenmann estimates that a search-based advertising platform could break even with about 7.5 percent of the global market. See Eisenmann, *supra* note 16.

"liquidity"—the volume of buyers and sellers that could reach mutually profitable trades—is well documented for exchanges.[37] Without enough liquidity markets are too thin and unsustainable.

In the case of search-ad platforms the advertisers are the buyers of access to Internet users while the searchers are selling that access through the ad search platform. More advertisers and more searches increase the likelihood of profitable matches given that advertisers and searchers are heterogeneous. It is useful to consider why in more detail.

Searchers obtain more relevant ads when there are more advertisers.[38] Suppose an individual is in Germany and needs a SIM card. She types in "Germany SIM cards" into Google. On September 26, 2007 she would have seen ten ads, of which eight are directly relevant to her query. If she typed the same query into the smaller MSN search engine on that same day she would have seen eight ads of which two are directly relevant to her query. That phenomenon is general: search-ad platforms with more advertisers will generally deliver more relevant ads to the searcher; that statement is particularly true for less common keyword combinations, for which there is a thinner advertising market. Since many searchers are looking to buy things, the larger platform is more valuable to them and they are therefore more likely to use the larger platform all else equal.[39] (In addition each searcher benefits from other searcher's queries because the search engine captures information that makes search results more relevant to subsequent searches.)

Advertisers also value more searchers. Consider an advertiser that earns $50 per unit on the sale of widgets. It has access to platform 1 and platform 2 where the first platform has 10 times as many searchers as the second. On average every click generates a sale 20 percent of the time. Suppose that it pays $0.50 per click to obtain the 3rd slot on each platform. Platform 1 sends 200 clicks per week generating 40 sales and platform 2 sends 20 clicks per week generating 4 sales. Then the advertiser earns $2000 in revenue per week from its campaign on platform 1 for which it pays $100 and earns a profit of $1900; it earns $200 in revenue per week from its campaign on platform 2,

[37] See, for example, Kenneth D. Garbade and William L. Silber, Structural Org*anization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk,* The Journal of Finance, Vol. 34, No. 3 (June 1979), pp.577-593; See also Nicholas Economides, *Network Economics with Application to Finance,* Financial Markets, Institutions & Instruments, Vol. 2, No. 5 (1993a), pp. 89-97.

[38] Many searchers are looking to buy something and therefore may value relevant advertising. It has been reported that about 40 percent of search queries involve potential commercial transactions. See Thomas Eisenmann, The Economics of Internet Advertising: Implications for the Google-DoubleClick Merger, Presentation for AEI-Brookings Joint Center, July 2007; see also Dai, et al, Detecting Online Commercial Intention, World Wide Web Conference, Edinburgh, Scotland, 2006.

[39] Search engines typically capture data on searches and accumulate this over time. (Maria Godoy, *Google Records Subpoena Raises Privacy Fears,* NPR.org, January 20, 2006. They use this information to improve the ability of the search engine to deliver relevant results. Some studies have found that there are not significant differences in the quality of search results across the major platforms despite the extreme differences in the number of searchers. Thus it would appear that direct network effects from search are limited. However, search and click-through histories enable search-ad platforms to estimate CTRs better and therefore gives rise to another potentially significant scale effect.

for which it pays $10 and earns a profit of $190. The advertiser therefore values access to the large platform more than it values access to the small platform, even though it does not value a click or a searcher on one more than the other.[40]

The existence of fixed costs, together with the difference in platform value documented above, may have an effect on the economics of the search-ad platforms, given the CPC pricing structure, especially for smaller platforms. Advertisers incur two costs of running campaigns that are independent of the number of clicks. First, they incur costs of setting up the platform, installing software, and learning how to use it.[41] Consequently, the advertiser must exceed a minimal volume of advertising (or more specifically, a minimal level of incremental profits) from this campaign before contracting with another search ad platform. Platform set-up costs discourage smaller advertisers from joining smaller platforms. As discussed above, many smaller advertisers advertise only on Google or only on Google and Yahoo.

Second, advertisers incur costs of running a campaign on keywords. They have to make decisions on the bids and monitor the performance of the campaign.[42] These tasks generally cannot be automated fully and therefore require humans. Thus, the advertiser must exceed a minimal volume of clicks on a campaign before mounting it on an ad platform that has been set up. To take the example above, if it cost $200 per week to monitor a campaign for widgets the advertiser would run the campaign on platform 1 but not platform 2. Therefore, campaign monitoring costs also discourage advertisers from mounting campaigns on platforms that have fewer search queries. As discussed above, even larger advertisers will tend to maintain smaller campaigns on the smaller search ad platforms.

These considerations lead to a positive feedback loop between the search and advertiser sides. To see this, consider starting in a situation in which two platforms have equal numbers of searchers and advertisers. Now suppose platform 1 has an exogenous increase of 10 percent in search traffic with platform 2 holding steady. That will result in some advertisers joining platform 1 that had previously found it unprofitable to join either platform, and in some advertisers mounting campaigns on platform 1 that they had previously decided not to mount on either platform.

Platform 1 now has more relevant ads for searchers. We would expect that some searchers would switch from platform 1 to platform 2. That in turn would increase the volume of advertising on platform 1. One could go through the same argument with an exogenous increase in advertis-

[40] This statement is true except for the situation in which it pays is maximum value per click of $10. Then it earns zero profit from either platform and is indifferent between them. While the second-price auction in theory is designed to get bidders to pay their maximum values in reality it is will not do so perfectly and we disregard this extreme situation.

[41] There is also an activation fee for the major search-ad platforms. Google and Microsoft each charge $5; Yahoo has no fee for the "Self Serve" version, but charges $199 for an assisted setup.

[42] Maintaining an advertising campaign requires choosing the right keywords and fine tuning them, modifying bids, selecting the best landing pages, revising ad text, and monitoring account statistics such as clicks, impressions, CTR, average CPC, average position, and conversion rate.

ing on the other platform. In both cases the effect of an advantage on one side becomes magnified as a result of the positive feedback effects.

Platform 1 obtains a further advantage as it obtains more advertisers. As a result of the keyword bidding system, an increase in advertisers may increase the bids on keywords. Consider first a situation in which we would expect the two platforms to secure identical CPCs. Platforms 1 and 2 have the same 20 bidders interested in 10 slots; the slots are as valuable on platform 1 as on platform 2; and the keyword auction is equally efficient in the sense of getting the bidders to reveal their highest values. In this case we would expect the auction to result in the same bids for the same slots. Now suppose that the number of advertisers on platform 1 increases exogenously by two advertisers, while the number of advertisers on platform 2 is unchanged. If the new advertisers are situated similarly to the existing advertisers (e.g., if they are all drawn from the same distribution), then it is likely that one of the new advertiser's optimal bid will place it in one of the top 10 slots. Suppose, for example, the new advertiser falls into the sixth slot. This has two main positive revenue effects for the platform. First, the new advertisers in slots 6 through 10 all have higher bids than before (because the new 10th place slot is taken by the advertiser previously in slot 9, and so on). With the higher bids come higher payments to the platform. The second effect comes from the new advertiser in slot 6 having a higher bid than the prior advertiser in slot 6. This higher bid increases the bid of every advertiser in slots 1 through 5, because their optimal bids depend (positively) on the level of the bidder below them.[43]

All else equal, positive feedback effects would lead one ad search platform to achieve a monopoly position.[44] The largest platform would always realize the highest CPC and provide the largest overall value to searchers and advertisers in the aggregate. The platform that provided the highest quality search engine for users and ad platform for advertisers would necessarily win the market. We would expect, though, that these positive feedback effects would diminish with the size of the platform. That is because the fixed costs become less important as the number of search queries expand, the value of additional bidders declines as the number of bidders increases, and more keywords have thick markets as the number of searchers expand.

3. The Role of Rent Extraction and Ad Placement

Another difference between search-ad platforms concerns the revenue-per-search (RPS) they obtain for comparable search traffic and keyword bids. Platform 2 could realize a higher RPS than platform 1 even though both have the same amount of search and advertiser demand. That could happen if platform 2 was better at extracting value from advertisers due to having a more efficient keyword

[43] See Varian, *supra* note 25.

[44] This is the classic tipping story for markets with indirect network effects. See Brian Arthur, *Increasing Returns and the New World of Business,* Harvard Business Review 74 (July–August 1996): 100–109; Michael Katz and Carl Shapiro, *Systems Competition and Network Effects*. Journal of Economic Perspectives 8 (Spring 1994): 93–115. Tipping does not occur in many real-world markets as discussed in David S. Evans and Richard Schmalensee, *The Industrial Organization of Markets Based on Two-Sided Platforms*, Competition Policy International, Spring 2007.

auction, or if platform 2 was better at predicting CTR, and therefore in maximizing revenue from page placement, or if platform 2 was able to deliver more relevant ads (and therefore better leads) that advertisers would pay more for. Together, these advantages would enable platform 2 to obtain a higher CPC for a given keyword and to obtain a higher CTR by offering more relevant ads to searchers. (We refer to this as an "RPS|S" advantage to denote that it is RPS controlling for size.)

Let us consider the implications of the RSP|Scale differences for the dynamic competition among platforms. One way for a smaller ad platform to catch up with a larger ad platform is to buy traffic. In fact, a substantial portion of Google's search ad traffic comes from third-party sites. AOL and Ask.com, the next largest search sites after Google, Yahoo, and MSN, use Google to supply paid search advertising results. Combined, they accounted for about 10 percent of search traffic in the United States among the top five search engines, compared to Google's 61 percent.[45] Google, Yahoo, and Microsoft have bid for a number of these exclusive search contracts with third-party publishers. Google has won most of the head-on competition for which public data is available. For example, of the top 10 web sites that offer sponsored search results, and are not operated by firms that also own search engines, 8 use Google to provide paid search results.[46]

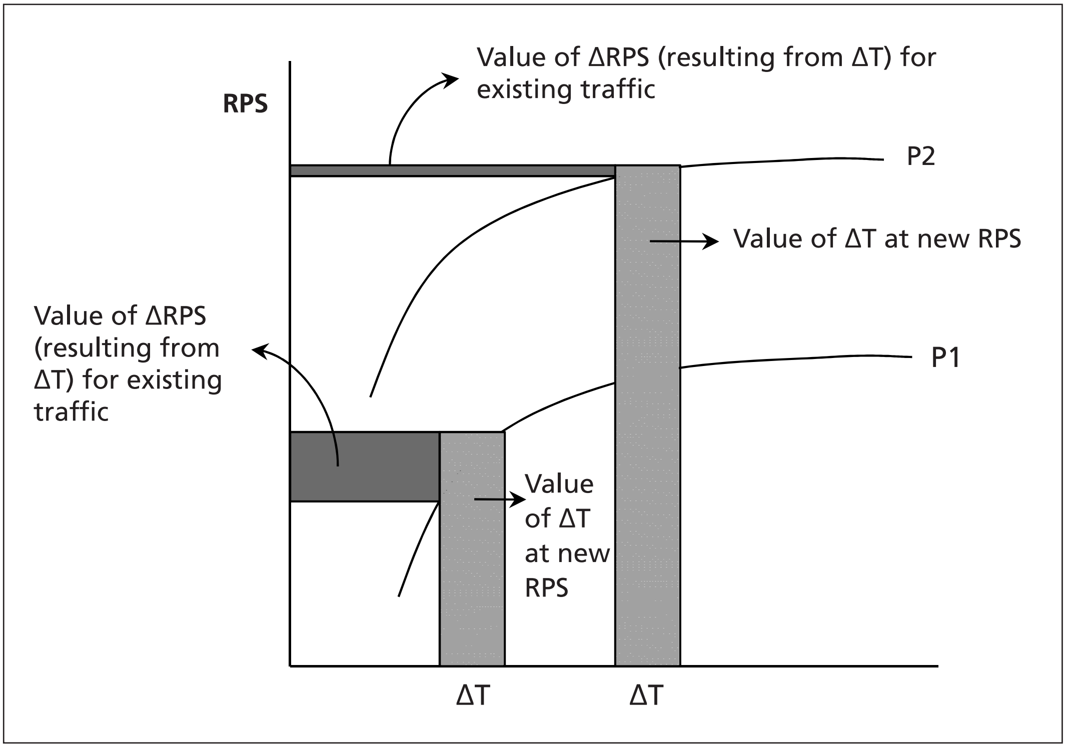
Consider the situation in which the only difference between two platforms is that platform 2 has more search traffic than platform 1. Moreover, there are decreasing returns to scale so the RPS increases but eventually flattens out with respect to scale (RPS'(S)>, RPS"(S)<0, and both eventually equal zero). Platform 2 has a higher RPS than Platform 1 but Platform 1 has a higher increase in RPS at its smaller size for any increase in search traffic. Then for sufficiently large increases in traffic Platform 1 will receive a larger increase in revenue than Platform 2 and it will have an incentive to outbid Platform 2. (For smaller increases in traffic, however, that increase taken alone, can be more valuable to the larger platform.) [47] More generally, Platform 1 could

[45] Discussions with industry analysts suggest that 25 percent or more of Google's search ad revenue comes from partner sites.

[46] Web sites owned by Google, Yahoo, Microsoft and Baidu were not considered. The website aol.com was considered, even though Google owns a 5 percent stake in AOL. Similarly, mapquest.com, which is owned by AOL, was considered. The website ebay.com uses Google for paid search ad listings outside of the United States but uses Yahoo in the United States. Go.com also uses Yahoo. Lycos.com uses Ask.com, which in turn uses Google for paid search ad listings. Amazon uses its A9 search site, which uses Microsoft for search. Download.com uses Search.com. Search.com is a "meta" search site and includes unpaid search results from Google, Ask.com, Microsoft and others. The paid search ad results on Search.com are frequently referenced to Google and we have counted it among the publishers using Google.

[47] For example, suppose the smaller platform had 10 searches with an RPS of 50 cents and the larger platform had 30 searches with an RPS of 100 cents and that an increase of 10 searches would increase the smaller platform's RPS to 70 cents and the larger platform's RPS to 105 cents. Even under these assumptions where the smaller platform gets a much larger increase in RPS, the increase in profits from getting the additional 5 searches is still larger for the large platform than the small one ($6.75 versus $5.50). In the long run, the smaller platform would also consider the aggregate benefits of accumulating additional traffic from multiple sources. The value of an increment in traffic comes not only from increasing its RPS on its existing traffic and the value of the new traffic, but also from the additional value that can result from future deals that increase its RPS on all of its traffic.

FIGURE 2 Traffic Volume



potentially equalize its RPS with platform 2 if it can buy enough search traffic to eliminate the difference between the two. That would seem possible if the search queries are expanding over time and the RPS scale effect eventually vanishes. (Note, also, that the smaller platform could buy traffic by paying users of Platform 2 to switch as opposed to entering into deals with publishers to carry its search tool bar.)

Now suppose that platform 2 receives a higher RPS at every level of search query traffic as shown in Figure 2. In this case, platform 2 has a bidding advantage over platform 1 at all sizes. Platform 1 can not achieve the same RPS by buying traffic. Moreover, if the difference between the curves is sufficiently large, platform 2 will continue to have an advantage even if the RPS curve flattens. Even if the smaller platform gets a greater increase in RPS from buying the same amount of traffic, it faces two disadvantages relative to the larger and more effective platform: (1) the value of the incremental traffic at the new RPS is still substantially lower than the other platform and (2) the value of the increase in RPS is applied to a smaller volume of existing traffic than for the larger platform. It would seem difficult for a smaller platform to catch up with a much larger platform that has also has a substantial RPS advantage at every given scale.

4. Factors That Work For and Against Winner-Take-All Competition

A platform that secures a lead generates more positive feedback effects thereby increasing its lead. Smaller platforms can catch the larger platform in this case only if they are able to secure more traffic and can eliminate the RSP|S difference with the larger platform. As a practical matter, once the larger platform has captured most of the market it is difficult for the smaller platform to catch up. Most traffic comes from an ad platform's own search site. To persuade users of the alternative platform to switch the second-placed platform has to offer sufficiently better search results to offset the superior advertising on the first-placed platform or pay them to induce them to switch. Likewise, to persuade advertisers to increase their use the second-placed platform despite having less traffic it has to subsidize or otherwise lower their fixed costs of managing campaigns. Neither effort appears to be impossible, but both seem daunting.

The existence of positive-feedback effects does not, however, lead most multi-sided markets to converge to monopoly as observed by Evans and Schmalensee (2007). Moreover, although firms secure quality advantages over each other in most real-world markets it does not appear that positive-feedback effects together with quality differences typically result in winner-take-all competition. Multiple competing platforms emerge when it is possible for smaller platforms to differentiate themselves from the leading platform, and from each other.

To see how such differentiation works in practice it is useful to consider three close analogies to search-ad platforms. Most traditional media markets have multiple players. That results from market segmentation in which media firms aggregate particular kinds of consumers and sell access to these consumers to advertisers for whom those consumers are particularly valuable. The magazine industry has taken this to the greatest extreme as inspection of any magazine stand demonstrates.[48] Financial markets have traditionally supported a few competing exchanges, although the move to electronic transactions has led to some consolidation. There is currently global competition among exchanges. Another two-sided market where we would expect monopolies is the yellow pages market. For most parts of the US however this is not the case. Although the local phone companies dominate some markets, private publishers have a significant presence in the most markets. In particular about 60 percent of submarkets have at least two yellow pages publishers, with roughly 25 percent of submarkets having more than 2 publishers.[49]

It remains to be seen whether ad search providers can differentiate enough to sustain a market structure that is not dominated by a single player. At this stage in the market evolution, it appears that search-based advertising is heading towards a single winner through some combination of positive feedback effects and RSP|S advantages. Google has increased its lead in search and advertisers each year since 2002. It has achieved a more than 75 percent share of search in 10 of the 15

[48] For example Future, plc is a UK company that focuses on specialized magazines. These include *Cross Stitcher, Guitarist, PC Gamer, Fast Car, Revolver, Mountain Biking UK, and Disney Girl.*

[49] Mark Rysman, *Competition Between Networks: A Study of the Market for Yellow Pages,* Review of Economic Studies (2004).

countries for which data were analyzed.[50,51] The fact that it currently receives a CPC that is more than 2 times larger than its nearest rival also suggests that it has a significant RPS advantage over its rivals that enables it to bid more than its rivals for incremental traffic.[52] In fact, Google has won most of the significant competitions to provide search toolbars or paid search listings on websites in the last couple of years, including a deal with AOL, one of the largest search sites of any significance besides Google, Yahoo and Microsoft, which operate their own search ad platforms.[53]

These facts would suggest that the positive feedback and RSP|S effects for Google are significant. To coexist, competing platforms would have to narrow the scale differences, narrow any efficiency differences that lead to a higher RPS for a given scale, and differentiate themselves to compensate for remaining scale and efficiency disadvantages. Competing platforms could also grow and potentially overtake Google by creating superior search engines that provide more relevant results for users and superior advertising engines that, among other things, deliver more profitable sales to advertisers. Such improvements in quality on the two sides of the market may explain how Google was able to overtake Yahoo which was the initial leader in search-based advertising.

IV. ADVERTISING ON PUBLISHER WEBSITES

Search results pages provide a large amount of advertising inventory which is primarily sold as text-based ads to advertisers. Many other web sites also supply advertising inventory as a result of having attracted traffic through content other than search results.[54] Worldwide, there were over 19,000 websites that reached more than 0.05 percent of the estimated 810 million unique visitors on the Internet and could therefore also supply advertising inventory.[55] These include traditional

---

[50] Source; comScore, "qSearch 2.0 Key Measures Report," November 2007.

[51] The fact that Yahoo lost its lead in search is of course not consistent with a pure positive feedback explanation for Google's success.

[52] See *supra* 37.

[53] Google also has deals with Fox, AOL, IAC and Dell. On August 7, 2006, Google and Fox Interactive Media entered into a multi-year contract making Google "the exclusive search and keyword targeted advertising sales provider for Fox Interactive Media." As part of the agreement Google will pay Fox at least $900 million so long as Fox is able to meet given traffic requirements and other commitments. See http://www.google.com/press/pressrel/ir_20060807.html. In December 2005, Google concluded a $1 billion deal with AOL which among other things would allow "AOL… to sell additional ads for its search engine also [which is] powered by Google." See http://www.news.com/AOL-to-stick-with-Google/2100-1030_3-5998600.html. Google and IAC (the parent company of Ask.com) extended their sponsored search and advertising agreement in a move which is worth an estimated $3.5 Billion to IAC over the next 5 years. See http://www.nytimes.com/2007/11/06/technology/06diller.html. Dell also recently (May 2006) concluded a deal with Google which "means millions of Dell computers will leave the factories with Google software already installed on them." See http://news.bbc.co.uk/2/hi/technology/5019416.stm.

[54] Search-engines are a special type of publishers. They use search result pages to attract viewers and then sell access to these viewers to advertisers.

[55] Source: comScore My Metrix report, November 2007.

media companies such as *The New York Times* and CNN that have established online extensions, new media companies such as You Tube, blogs such as technorati.com,[56] social network sites such as Facebook; and web portals that operate their own ad platforms including MSN and Yahoo. The ad inventory on these websites is often called "display" and often looks much like one sees in print media. Unlike the suppliers of ad inventory on search results pages, the suppliers of display advertising inventory usually charge based on the cost-per-thousand viewers (CPM); that pricing model may change as discussed below.

The display advertising business has a complex business ecosystem. Unlike search which is based on an integrated two-sided platform that sits between the viewer and the advertiser, display advertising consists of several different platform models. These platform models differ according to the extent to which intermediation between the publisher and the advertiser is integrated into the publisher that supplies the inventory and the extent to which software tools that link advertisers and publishers are integrated into either the publisher that supplies the inventory or the intermediary between the publisher and the advertiser. The provision of advertising to viewers is therefore often provided by two or more interconnecting two-sided platforms. Part A provides an overview of the display advertising business. Part B describes the current market structure, including the role of different two-sided platforms. Part C discusses ways in which this industry, which is currently going through significant changes, may evolve.

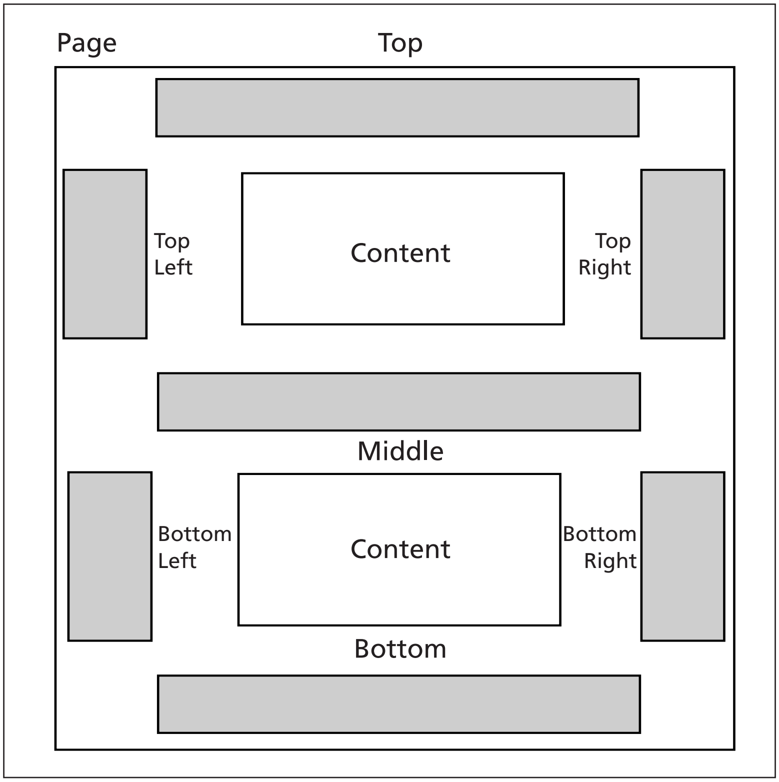
A. Overview of the Online Display Advertising Business

Web publishers create advertising inventory by designing their web pages, which are written in HTML, to accept graphical, text, or video ads in various portions of the page.[57] These portions of the page are reserved for ads and usually include code that permits the insertion of advertising on a real-time basis from various sources. Because of the necessary coding most websites do not change the space made available for ads frequently unlike newspapers, which can readily modify the layout for each print run. If they do not have an ad available to insert in a space they will often use the space for self-promotion. Figure 3 shows a typical layout of a page. Ad space is sold based on the length and width of the space in pixels; typical types of ads include banner ads at the top, skyscraper ads on the sides, and rectangle ads in various locations. As with traditional print advertising, some space is more valuable than others (top right vs. bottom left), some viewers are more valuable than others (18-49 vs. 50+), and some sites are more valuable than others (finance vs. entertainment news) in part because of the type of viewers they attract and in part because of what those viewers might be predisposed to purchase.

[56] David S. Evans, Economics of the Blogosphere, The Catalyst Code, August 6, 2007, http://www.thecatalystcode.com/theconversation/blog/2007/08/06/economics-of-the-blogosphere/.

[57] Web pages can also include languages such as JavaScript that generate HTML code when a page is called.

FIGURE 3 Relative positioning of ads on a web page.



The key technological difference between traditional display advertising and online display advertising involves the ability to target ads to particular consumers in real time. In the traditional advertising industry targeting is primarily accomplished through the development of media that use content to attract particular socioeconomic and demographic groups (from broad groups such as people between 29 and 55 to narrow groups such as fly fishermen). Traditional advertisements are seen by all views of the media containing the ad. Online advertising makes it possible to serve an advertisement specifically for the viewer of the web page based on knowledge of the viewer's socioeconomic characteristics or even what sorts of products the viewer is considering purchasing.

Online advertisers and publishers currently engage in targeting to a relatively modest degree. An advertiser will buy the front page of Yahoo, which is not that different from buying advertising on a television show—the advertiser is interested in reaching those audiences but is not targeting beyond that. Many of the investments currently made in online advertising (including the purchases of ad networks and ad serving and behavioral targeting providers), however, reflect a belief that more sophisticated targeting will become increasingly important.

Unlike traditional advertising, online ads can use the IP address of the computer that has signed on to the website to focus on narrow geographic areas including ones with particular socioeconomic characteristics. Publishers also have the ability to determine what ads viewers have seen recently on their sites and limit duplicate messages. They can also determine what content the user has recently been interested in—for example, an auto advertiser may be interested in reaching viewers who have recently looked at the auto reviews, but with online advertising these viewers can be shown ad not only on the auto section but on other content pages they may be viewing. Viewers for whom the publisher has registration or other identifying information offer further targeting opportunities. Contextual advertising platforms, as discussed below, provide some additional targeting by considering whether certain keywords such as "luxury hotels London" occur on the web page. It is expected, however, that online advertising will increasingly make greater use of viewer data—whether actual or inferred—to better target more ads to viewers and thereby make online advertising make effective for advertisers. We discuss such "behavioral targeting" below in examining future developments in the online advertising business.

Like all sellers and buyers, publishers and advertisers require ways to identify optimal trading opportunities and to establish transaction prices. There are two major ways that this "intermediation" occurs. First, it can occur directly through bilateral exchanges between publishers and advertisers. eBay, for example, may sell Nokia the right to present an advertisement in a particular spot to viewers with specific characteristics by having its sales agents deal directly with Nokia's buying agents. Second, it can occur *indirectly* through multilateral exchanges between publishers and advertisers using advertising networks. Hearst Publishing may sell ValueClick—an advertising network—advertising inventory from its various online newspaper and magazine properties. ValueClick may then sell advertisers access to particular kinds of viewers by pooling relevant advertising inventory from Hearst and other publishers. In this case, the advertiser buys access to a type of viewer—for example "fashion conscious young women in upscale locations"—but is not necessarily guaranteed that it will reach these women on Hearst's *Cosmopolitan* website.

The extent to which advertisers and publishers use direct and indirect methods of distribution for advertising varies. Smaller publishers and advertisers typically rely only on indirect methods because it is not economical to carry the cost of salespeople and purchasing agents. Larger publishers and advertisers—which account for the preponderance of online display advertising sales—often use both direct and indirect methods. Larger publishers use direct methods to sell their premium space—such as the space on the first web page and high quality space on some of the interior web pages—and this space is then said to be "reserved" for an advertiser. They also use indirect methods to lower quality space such as the bottom space on interior pages, or advertising inventory that has not been reserved either because there was an unexpected spike in viewership which increased the supply of inventory, or because they did not sell all of their advertising inventory at the prices sought in direct sales. Advertising inventory that has not been reserved is sometimes called "remnant". We discuss the relative pricing and use of directly and indirectly sold online advertisements below.

Advertisers and publishers rely on various services which they either provide themselves or purchase from others. Large advertisers and advertising agencies often license an advertising package as aQuantive's AdManager.[58] They typically manage advertising on hundreds of websites and across numerous products using many methods of online advertising. The advertising software helps them manage these various campaigns and instructs the server to deliver advertising copy downstream into available inventory.

Large publishers usually license a publisher tool such as DoubleClick's DART for Publishers (DFP).[59,60] Publisher tools are highly sophisticated software programs. They are usually integrated into many other aspects of the website technology and business practices. The publisher hard codes links to the publisher tool to fill the ad space for which it wants to use the management, reporting, and serving capabilities of the publisher tool. When a viewer clicks on a web page the publisher tool makes a decision, based on preprogrammed rules, as to which ad to present to the viewer. The publisher tool checks whether the particular ad space that the viewer is about to see has been "reserved", and if not, whether there is an ad network that can fill that inventory space. The publisher tool then retrieves the chosen ad, or directs the advertiser or ad network's server to retrieve the ad, which is then shown to the user. This entire process happens in the blink of an eye. A few very large publishers rely on publisher tools they have developed themselves rather than licensing a publisher tool from a third party.

Publishers also use a variety of advertising networks for indirectly sold advertisements as we discuss below.

B. Market Structure for Advertising on Publisher Websites

The online display advertising business consists of several types of two-sided platform businesses. Some platforms integrate features of the other platforms as we will see below.

- Publishers supply content on their websites that attracts viewers and sell the resulting advertising inventory to advertisers directly or indirectly.
- Advertising networks serve as intermediaries between publishers and advertisers by aggregating the demand and supply of advertising inventory. (Other intermediaries such as third-party sales representative also operate two-sided intermediation platforms.)
- Advertising software tool companies usually supply both software tools for publishers and software tools for advertisers; they at least try to generate positive feedbacks between the publisher and advertiser side.

Table 5 lists the top two firms in each of these categories and estimates of their revenues.

[58] See http://www.atlassolutions.com/services.aspx and http://www.doubleclick.com/products/dfa/index.aspx.
[59] See http://www.doubleclick.com/products/dfp/index.aspx.
[60] A handful of mega-large web publishers such as MSN have their own proprietary tools but most others use a third-party tool.

TABLE 5 Leading Publishers, Advertising Networks and Tool Providers

	2006 Revenue (M)
Publishers (Display Advertising)	
Yahoo	$1,158
Time Warner / AOL	$1106
Microsoft Network	$939
New York Times / About.com	$240
Advertising Networks	
Google	~$1,500
Advertising.com	$455
ValueClick	$383
Right Media	$150
Tool Providers	
DoubleClick	~$300
24/7 Real Media	$200
Atlas	$122

Source: Publishers' revenues are based on Digital Marketing and Media Fact Pack: A Supplement to Advertising Age, Casale Media Inc., April 23, 2007; Google's revenue is based on Google's 2006 10-K report; Advertising.com's revenue is from Time Warner's 2006 10-K report (Advertising.com is a wholly owned subsidiary, and generates one third of AOL's advertising revenue); ValueClick's revenue comes from ValueClick's 2006 10-K report; Right Media's revenue comes from The Quest for the Perfect Online Ad, Paul Sloan, CNNMoney.com, April 3, 2007, Article available at: http://money.cnn.com/magazines/business2/business2_archive/2007/03/01/8401043/index.htm?postversion=2007040 309; DoubleClick's revenue comes from Google Buys an Online Ad Firm for $3.1 Billion, Louise Story and Miguel Helft, New York Times, April 14, 2007. Article available at http://www.nytimes.com/2007/04/14/technology/14deal.html?ex=1334203200&en=d94eb7f788b32db5&ei=5090&par tner=rssuserland&emc=rss; 24/7 Real Media's revenue is based on 24/7 Real Media's 2006 10-K report; Atlas' revenue is based on aQuantive's 2006 10-K report.

Notes:	i) Rankings and revenues of publishers are for the United States only. All other rankings and revenues are based on world-wide data.
	ii) The Advertising Age figures for publisher revenue are industry estimates.
	iii) Google does not report contextual revenue publicly. The given figure is a rough estimate based on discussions with industry experts.
	iv) The revenue number for ValueClick is for the company's Media segment, which includes Display Advertising, Lead Generation Marketing, Email Marketing and E-commerce, so the display ad revenue may be somewhat overstated.
	v) The revenue figures for DoubleClick and 24/7 RealMedia are for the respective companies' entire revenue, including but not necessarily restricted to revenue from the provision of software tools.
	vi) Atlas' revenue figure is from aQuantive's 'Digital Marketing Technologies (DMT)' segment. In its annual report, aQuantive states that "[o]ur DMT segment consists of Atlas, a provider of digital marketing technologies and expertise."

Online publishers in turn come in three variations.

- Large integrated online publishers have their own publisher software tools for operating their websites. In addition to their direct sales forces, they also have advertising networks that sell their non-reserved inventory to advertisers.
- Large nonintegrated publishers have their own direct sales forces. However, they usually license a third-party publisher tool and work with one or more third-party advertising networks to sell their non-reserved inventory to advertisers.
- Small nonintegrated publishers typically do not have their own direct sales forces. They usually rely on an advertising network that has its own publisher tools. Once they paste a direction to that advertising network in their advertising space they need do nothing more.

Detailed data on the sales of these three types of publishers is not available. However, discussions with industry experts suggest that the substantial majority of display ad revenue are for ads placed on the first two types of publishers. The largest publishers— , AOL, MSN and Yahoo—are integrated, fall under the first category, and collectively account for ad revenue that is likely on the same order of magnitude as the nonintegrated publishers.

There are two major kinds of advertising networks.

- Contextual ad networks use the same software that they use for managing their search-based advertising campaigns to manage their contextual-based advertising campaigns. Advertisers use this software to bid on keywords on the publishers that rely on the contextual network. The contextual ad network then insert ads provided by the advertiser on those websites based on the advertiser's bid and the content of the sites on its network.[61] Advertisers also have some ability to select inventory based on the website or based on various characteristics of users.
- Most online networks do not use contextual targeting. These standard online advertising networks specialize in aggregating advertising inventory from a variety of publishers that corresponds to some group for which there is demand from advertisers. They then sell advertisers the ability to present their ads to some number of viewers in that group without necessarily guaranteeing that the ads will be seen on any particular site. As mentioned below some of these ad networks are using behavioral targeting techniques to deliver more relevant ads for viewers and advertisers.

Table 5 lists the leading advertising networks worldwide. Google operates by far the largest contextual advertising network and has the largest advertising network overall, with ad revenues

[61] The contextual platforms also used other characteristics for targeting in addition to content.

that are likely three times those of its next closest competitor in 2006. Other leading firms include Advertising.com with revenue of $455 million is 2006, and ValueClick with revenue of $383 million in 2006. A number of even smaller competitors, such as Tribal Fusion, BlueLithium, Quigo and Tacoda are not publicly traded and financial data on their ad revenues are not available.

Finally, the leading software tools providers make tools for both publishers and advertisers even though these tools are in some sense very different software. They have at least attempted to persuade publishers and advertisers that they would benefit from their tools because the other side of the market is using those tools. On the DART for Publishers web page, for example, DoubleClick states, "[o]ur position at the center of the digital media landscape means we can deliver unparalleled buyer/seller integration, including zero discrepancies on ads trafficked from DoubleClick's DART for Advertisers solution to its DART for Publishers solution. This puts money directly into your pocket and simplifies billing and reconciliation."[62] A key development in 2007 was that the leading software tool providers were purchased by other platforms. Google bought DoubleClick, Microsoft bought aQuantive, and WPP—a large advertising agency—bought 24/7 RealMedia. DoubleClick is generally regarded as the largest tools provider,[63] especially for major publishers. Table 5 lists the leading tools providers worldwide. In 2006, DoubleClick had revenue of about $300 million. Tools providers not listed in the table include MediaPlex and ValueClick.

There is a close relationship between the form of intermediation used by publishers and their use of publisher tools and advertising networks. Publishers that want to sell directly must have a third-party publisher tool or provide their own software. Publishers that want to sell some space indirectly must use an advertising network. Comprehensive data are not available on the portion of directly and indirectly sold advertising inventory. Advertisers and publishers I have talked to, however, seem to agree that somewhere around 70 percent of advertising revenue for large publishers is the result of direct sales and the remaining 30 percent of advertising revenue the result of indirect sales through contextual or traditional advertising networks. Because indirectly sold ad inventory is much cheaper than directly sold ad inventory, the advertising networks account for a higher percentage—perhaps around 50 percent—based on the number of "impressions" (the viewers that see an ad). However, some individuals in the industry put the direct sales figure as high as 95 percent of revenue for large publishers.

C. Prices and Costs for Online Display Advertising

Most online display advertising inventory, on a revenue basis, is bought and sold on the basis of CPM. The prices paid by advertisers varied from less than $1 to more than $100 as of early 2007. Where the price falls in this range depends on the interplay of the location of the advertising inventory on the web page and the website; the amount of space measured in pixels; the characteristics

[62] See http://www.doubleclick.com/products/dfp/index.aspx.
[63] See http://www.redherring.com/Home/pages/print/posts/?bid=28262ee6-7dba-4f55-bb53-cd1a4e1d52c0&mode=Full.

of the viewers; and the type of intermediation used. As is common in business-to-business transactions large publishers often engage in individual negotiations with advertisers and discount prices off of a rate card. The contextual ad networks typically charge advertisers on a CPC basis which is determined from the keyword bidding process. Not surprisingly, large publishers receive payments that are in the same range as for advertising networks for the same amount of space (three or four contextual ads typically appear in the same space), on the order of $2 CPM.

The publishers have to pay for intermediation services when they are not using a direct sales force to sell the ad. Google, the leading contextual network, is believed to keep about 15-20 percent of advertiser revenues.[64] The standard advertising networks typically keep 20 percent or more, although the negotiations are highly individualized.[65] The tool providers also impose a transaction cost on the buying and selling of advertising inventory. In all cases the advertising networks keep a portion of the revenue they receive from advertisers as a commission. For directly sold ads the publisher tools provider will usually charge about 5 cents per thousand impressions served and the advertiser tool provider about 7-8 cents per thousand impressions.

D. Market Evolution

For large publishers the online display advertising business does not look much different than the traditional advertising business but for the fact that space can be customized for each viewer. A large fraction of the revenues come from old-fashioned selling of ads by hand. And even for space that is not reserved the ad networks are not much different economic beasts than the American ad agency created in 1841 that bought space and sold it at a discount to publishers. Several developments could dramatically alter this business.

The first involves the use of computerized matching, clearing, and settlement of transactions between advertisers and publishers. The advertising networks currently do much of this work through human effort—they call publishers and advertisers. Some exchanges have been launched, which provided an electronic exchange between advertisers and publishers. A few exchanges also connect advertising networks to facilitate the trading of excess supply and demand between them. These exchanges face several challenges. Online advertising space is highly heterogeneous. Matching supply and demand is therefore a technically challenging exercise. Moreover, these exchanges would face the usual challenge of securing sufficient liquidity to have thick enough markets for potentially

64 Across all of Google's partner traffic, including both contextual and search, in the 12 months from Q4 2006 to Q3 2007, Google retained a 16 percent share. This share likely varied by type of traffic and publisher. See Google 2006 10-K Form, Google 10-Q form for the quarterly period ended September 30, 2007. To the extent the percentage retained differs significantly between search and contextual, then this figure may over or understate.

65 In discussions with industry experts, I have heard the share retained by the ad network may be as high as 50 percent in some cases. Typically, the ad network offers the publisher a fixed CPM guarantee and retains whatever increment it can sell the inventory for.

narrowly defined inventory.[66] However, they also have the opportunity to reduce what appear to be rather high transaction costs for buying and selling advertising inventory. Based on the figures reported above the transactions costs incurred by large publishers and advertisers is on the order of 20 percent or more of the sales price after the ad networks and publisher tool providers are taken into account. That is much higher than intermediation costs in other industries—0.1-1.7 percent for financial exchanges[67] and as little as 1-2 percent for some real estate transactions.[68]

The second development increases the value of electronic exchanges and is therefore could help these exchanges achieve a critical mass of liquidity. Data-intensive behavioral targeting can significantly increase the likelihood that an ad will reach a consumer to whom that ad is relevant and who is likely to translate into a sale for the advertisers. As this technology improves the value of selling ads through automated processes rather than through direct selling would increase. This sophisticated targeting faces at least two challenges. The first is that it is technologically difficult to process the relevant data and serve the advertisement so that it is part of the web page when the viewer first looks. The second is that behavioral targeting relies on extensive personal information the use of which has drawn increased scrutiny by regulators and the media.[69] Online advertising companies currently collect and store significant data through the insertion of cookies on personal computers and other techniques. It is possible that these activities will be curtailed.

The market structure that would result from these developments would have one or more advertising platforms sitting between publishers and advertisers. Trading platforms typically have significant scale economies and positive feedback effects. That results in a small number of viable players. In the case of financial exchanges the globalization of markets combined with the development of electronic trading has led to consolidations and increased concentration. Several factors suggest that there would be stronger forces leading to the emergence of a single advertising platform than has been the case for financial exchanges. First, behavioral targeting requires the use of histori-

---

[66] These are B2B exchanges. We know from the dramatic collapse of B2Bs in the early 2000s that obtaining liquidity, especially from the supply side (publishers in the case of online advertising) is the most serious challenge. See Silicon.com, *B2B exchanges - What happens now?,* January 23, 2002; Hazel Ward, *B2B exchanges fail to deliver*, Computer Weekly, November 29, 2001; Bob Tedeschi, *Investing; How Killer B-to-B's Went Into A Tailspin,* New York Times, May 7, 2000.

[67] Data are for 2004. *See Competition in the EU securities trading and post-trading-Issue paper,* European Commission, 2006.

[68] Data are for 2002-04. See *Estate agency market in England and Wales,* OFT, 2004.

[69] The FTC has heard privacy cases against several web properties, including the online advertiser ValueClick, which was ordered in March 2008 to pay a fine of $2.9 million "to settle Federal Trade Commission charges that its advertising claims and e-mails were deceptive and violated federal law." See http://www.ftc.gov/opa/2008/03/vc.shtm. The FTC has also published a report on privacy issues: *Protecting Consumers in the Next Tech-ade: A Report by the Staff of the Federal Trade Commission,* March 2008, available at http://www.ftc.gov/os/2008/03/P064101tech.pdf. Further, there have been moves towards legislation in the states of New York and Connecticut that would restrict the use of personal information for online advertising. See http://www.nytimes.com/2008/03/20/business/media/20adco.html.

cal information on consumers often pulled from their past behavior as observed by the advertising platform. Larger platforms could secure advantages that would be hard for smaller platforms to replicate. Second, behavioral targeting may have some of the long-tail effects that have given large search platforms advantages. As the supply and demand is sliced more finely, larger platforms have thicker markets in every niche. Third, effective behavioral targeting is likely to require significant technical advances as well significant computer capacity. The IP rights in particular could establish barriers to smaller players.

Winner-take-all competition stories, however, tend to work better in theory than in practice in many industries. While it seems implausible that there will be many commercially significant advertising platforms, two factors identified in the multi-sided platform literature could lead to several viable platforms. The first is whether it is possible for advertising platforms to differentiate themselves sufficiently from their rivals from the standpoint of advertisers or publishers while maintaining sufficient liquidity. The second concerns the extent to which advertisers or publishers can multi-home. There would not appear to be any obstacles to advertisers using multiple advertising platforms. Publishers currently rely on their publisher tools to multi-home with multiple alternative channels. If publisher tools become integrated into the advertising platform the ability to multi-home is likely to decline, thereby reducing competition (except for exclusives) at the advertiser level.

The development of advertising platforms for trading ads and eyeballs would alter market structure in the publishing industry as well. Most online and offline media companies are vertically integrated into selling advertising. They have their own sales forces which, for the larger publishers, sell the bulk of advertising. As advertising platforms evolve, publishers may specialize in attracting content and leave the advertising business to the platforms. Many small publishers now do just that.

V. CONCLUDING REMARKS

Internet-based technologies are revolutionizing the stodgy $625 billion global advertising industry. Advertisers once had only crude methods available for targeting their ads to consumers who were likely to buy their products. That was done mainly by selecting advertising media—such as particular television shows or magazines—that specialized in the relevant audiences. Advertisers also had little information on who actually watched their ads and what activities followed. The Internet has changed that by allowing advertisers to target specific individuals and paying only when those individuals click on the ads. More sophisticated technologies are beginning to track not only whether individuals clicked on an ad but whether that actually translated into a sale. Search-based advertising has developed the most advanced methods for targeting consumers and charging for results. Display based advertising is catching up.

These new technologies are critically important for understanding the evolution of the advertising business because it is expected that more advertising will take place online. Web-based businesses are providing more content and services that people want to spend time viewing. The number of Internet users worldwide increased from 741 million in December 2006[70] to 816 million in December 2007.[71] Younger people in particular are spending more time on social networking sites such as Facebook and online gaming sites such as Pogo.com.

In addition, more of the devices that people use for consuming content will have an Internet connection in the future. More than 1 billion PCs will be in use worldwide by the end of 2008.[72] There are 3 billion mobile phone subscribers worldwide and nearly 2 billion smart phones in use.[73] An increasing portion of those mobile phones will have internet connections and browsers as does the iPhone. It is expected that many televisions will have Internet connections as well.[74] Other devices may have internet connections as well. Amazon's e-book reader has an always-on wireless feature that would enable it to deliver targeted ads as well.

Finally, many of the advertising-supported web-based platforms have encouraged developers to write applications that work with their platforms. These applications may also carry advertising. For example, Google has encouraged developers to write applications using Google Maps but has reserved the right to insert advertisements into those applications.[75]

These changes are important for understanding the current structure and evolution of the online advertising business. They are also critical for numerous public policy issues that have emerged. Will a single ad platform emerge or will several remain viable? What are the consequences of alternative market structures for ad platforms for a web economy that is increasingly based on selling eyeballs to advertisers? Data is central to these ad platforms. Historically, communication providers such as the telephone companies have been highly respectful of privacy rights. The ad platforms have business models that are based on collecting and hoarding highly personal data from individuals. The implications of that have not been fully thought through by either consumers or policymakers.

---

[70] comScore MyMetrix Custom Media Trend Report, June 2007.
[71] comScore MyMetrix Key Measures Report, December 2007.
[72] See http://www.forrester.com/Research/Document/Excerpt/0,7211,42496,00.html
[73] Associated Press, *Yahoo revs up mobile phone - Plan hopes to attract new users, more ads,* The Star-Ledger, January 7, 2008; Alan Duerden, *CIOs On The Back Foot Over Handheld Device Security,* The Banker, October 1, 2007.
[74] See http://tvdecoder.blogs.nytimes.com/2008/04/16/tv-but-with-connections/.
[75] See http://www.google.com/accounts/TOS

CHAPTER

8

The Online Advertising Industry: Economics, Evolution, and Privacy

David S. Evans

ABSTRACT

Online advertising accounts for almost 9 percent of all advertising in the United States. This share is expected to increase as more media is consumed over the internet and as more advertisers shift spending to online technologies. The expansion of internet-based advertising is transforming the advertising business by providing more efficient methods of matching advertisers and consumers and is transforming the media business by providing a source of revenue for online media firms that compete with traditional media firms. The precipitous decline of the newspaper industry is one manifestation of the symbiotic relationship between online content and online advertising. Online-advertising is provided by a series of interlocking multi-sided platforms (also known as two-sided markets) that facilitate the matching of advertisers and consumers. These intermediaries increasingly make use of detailed individual data, predictive methods, and matching algorithms to create more efficient matches between consumers and advertisers. Some of their methods raise public policy issues that require balancing providing consumers more valuable advertising against the possible loss of valuable privacy.

Advertising delivered over the internet—"online advertising"—has become a significant source of revenue for web-based businesses. Fifty-six of the top 100 websites based on page views in February 2008 presented advertising; these 56 accounted for 86 percent of the total page views for these 100 sites.[1] Twenty-six of these 56 sites, accounting for 77 percent of all page views for the top 100 sites, likely earn most of their revenue from selling advertising. Advertising is a significant source of revenue for many of the smaller sites including ones such as blogs that occupy the long-tail of the web. Online advertising is also central to the $34 billion e-commerce economy (U.S. Department

[1] The top websites were based on ComScore MediaMetrix rankings. A researcher reviewed each website to determine whether it presented advertising on the home page or immediate branches of the home page. They determined whether advertising was the most significant source of revenue based on considering other apparent sources of revenue and the business model followed by the site. The appendix provides the details. The top 100 websites accounted for 41 percent of all page views on the web according to ComScore.

of Commerce, 2008), which is becoming an ever larger portion of overall commerce. Web-based sellers use online advertising to drive consumers directly to their sites where they can browse for goods and services and purchase them with a few clicks. Online advertising accounted for 8.8 percent of all advertising in the United States in 2008 (Hallerman, 2008). That share is expected to grow as people spend more time with new online media such as social networking and sites that offer user-generated content; as more traditional media such as television is transmitted over the internet; and as more advertising is delivered to browser-enabled mobile phones.[2]

Internet-based advertising is the source of a "gale of creative destruction" (in the words of Schumpeter, 1942) that is sweeping across the advertising and media landscape, especially in the United States. Newspapers, particularly, are losing readers and advertisers to web media supported by online advertising. That has lead to a downward spiral as indirect network effects work in reverse. The market caps of the major publicly traded newspaper businesses in the United States declined by 42 percent between January 2004 and August 2008, compared to a 15.6 percent increase for the Dow Jones Industrial average over that same time period (Blodget, 2008). With the additional pressure of the financial crisis, newspapers in several major cities including Denver and Seattle have closed down and others including the New York Times are in distress. More generally, online advertising is disrupting all aspects of the global advertising industry, which had estimated revenues of $625 billion in 2007 (Minton, 2007), from how creative work is done, to how advertising campaigns are run, and to how advertising is bought and sold.

Online advertising methods are, arguably, leading to significant reductions in transactions costs between merchants and consumers. The methods enable merchants to deliver information that is targeted to those consumers who value the information the most and are most likely to act on it. An oft-quoted line in the advertising business states ruefully: "Half the money I spend on advertising is wasted. The only trouble is I don't know which half." The new techniques replace a sledgehammer with a scalpel. In doing so, they collect and analyze detailed information about how people use their computers—raising difficult issues concerning the expectation of privacy and the regulation of the online advertising industry.

No research has yet examined the value of the productivity improvements created by online advertising technology but they appear significant. Consider a business that sells saltwater fishing rods to people who enjoy fly fishing. The traditional approach to matching this buyer and seller involved the creation of a magazine, such as *FlyFisherman*, with content that attracts the relevant people. In contrast, the online approach relies on a variety of techniques to match an advertising message to a consumer. A search engine indexes web results that are relevant

[2] Moreover, a consortium of cable television providers are developing methods for delivering targeted advertising to cable subscribers over their networks; although the advertisements are delivered over the closed cable networks rather than over the internet the targeting methods and business models are similar to those used for online advertising. Tim Arango, Cable Firms Join Forces to Attract Focused Ads, NY Times, Mar. 10, 2008.

to a consumer who types in the phrase “saltwater fishing rod,” and with this information, the search engine can sell ads to sellers of saltwater fishing rods. Contextual advertising on web pages could do the same thing. A consumer who visits a blog for fly fishermen could be presented with an advertisement. Developing behavioral targeting techniques, discussed below, can also identify individuals who are interested in fly fishing and determine whether they are looking around the web for information that would suggest they might be in the market for a saltwater rod.

This chapter presents the evolution of the online advertising business. It examines the supply of online advertising “inventory,” which equals the space times the views of that space; the demand for that inventory; and intermediaries that operate between the sell and buy sides. It also explores some of the key developments such as behavioral targeting for matching advertising messages to consumers and considers some economic aspects of the privacy issues that these technologies raise. Aside from providing a survey of an important new segment of advertising this essay suggests, at various points, that many of the interesting questions, and economic puzzles, about the advertising industry—offline and online—remain to be addressed. For example, online advertising is a “two-sided market” (Rochet and Tirole, 2003; Anderson et al., 2005), as is advertising generally. Intermediaries operate platforms that facilitate advertisers and consumers connecting with each other. Innovative intermediaries operate exchanges and face the critical liquidity issues discussed in the market microstructure literature (O’Hara, 1998; also see Evans and Schmalensee, 2009).

I. EVOLUTION OF ONLINE ADVERTISING

Online advertising started in 1994 when HotWired, a web magazine, sold a banner ad to AT&T and displayed it on their web page (Kaye and Medoff, 2001). The ad was sold based on the number of “impressions”— individuals who saw the ad—which was the model followed by most traditional media for this sort of brand advertising. Many web ads were subsequently sold based on “cost per mille,” which is advertising terminology for cost per 1000 viewers of the advertisement and often referred to as CPM. Paying by number of viewers remained the norm until Procter & Gamble negotiated a deal with Yahoo! in 1996 which compensated the web portal for ads based on the “cost-per-click” (commonly known as CPC). Yahoo! was paid only when a user clicked on the ad; this was the web-version of the direct response method commonly used by advertisers for things such as mail and telephone solicitations. However, we will see that most “display ads” on websites—the ads that look like those in newspapers and magazines— were still sold based on thousands of views as of 2008.

The exploding supply of web pages led to the birth, in 1994, of several search engines that also sold advertising to make money. At first, they sold banner ads on a cost-per-mille basis—that is, based on how many people saw the ad. However, that approach led to a conflict for the search engine between helping people find things quickly and keeping eyeballs trained on the site to make money. The search engines later moved to the cost-per-click model. GoTo.com—which is now owned by Yahoo!—introduced many of the key technological and business model innovations in the next three years ("GoTo to Overture," 2005; "History of Pay Per Click, 2007). These included adopting the cost-per-click approach to pricing and the use of auctions to allocate the advertising spots on the page showing results of the search.

During this same period, fairly traditional methods of advertising were mimicked on the web. These included web versions of business directories similar to the yellow pages such as yellowpages.com; web versions of newspaper classified ads such as Craigslist; and web versions of direct mail and telephone marketing such as CheetahMail. These web-based vehicles were charged for in ways that were similar to their traditional counterparts with the exception of Craigslist, which provided a significant amount of free advertising to consumers. The remainder of this chapter does not discuss web-based directories or email advertising because they do not raise particularly novel issues.

Online advertising revenue has increased steadily over time in absolute terms and as a fraction of all advertising revenue. Consistent figures are available since 2000. They show that U.S. online advertising has increased from $8.1 billion in 2000 to $21.2 billion in 2007 and from 3.2 percent of all advertising to 8.8 percent over that time period (based on Interactive Advertising Bureau Press Releases 2000-2007). The relative mix of online advertising has also changed. Table 1 shows the evolution of various online advertising formats from 2000 until the first half of 2008. In 2008, search and display-related ads were the leading advertising formats with 44 per cent and 21 per cent share of the total revenue, respectively. However, ads that are linked to the results of a keyword search have grown explosively from 1 percent of online advertising revenue in 2000 to 44 percent in 2008, while display ads dropped from 48 percent of total revenue in 2000 to 21 percent of total revenue by the first half of 2008.

In 2009, fifteen years after its birth, the online advertising industry remains in considerable flux. The delivery of online advertising exhibits rapid technological change. At the same time, new economic structures are emerging and business relationships among the key players are changing. For example, in 2008, Google completed its acquisition of DoubleClick, which was a major technology provider, and Microsoft entered into negotiations, ultimately aborted, to buy Yahoo! Nevertheless, certain features of the "online advertising ecosystem" have become clear and provide a useful framework for this discussion.

On one side of the business are advertisers that want to reach consumers and consider online advertising as a possible way to do that. On the other side are consumers who may or may not be receptive to receiving advertising messages. In between are various intermediaries. The fully integrated intermediaries touch consumers and advertisers directly. The search-based advertising platforms (Evans, 2008) are examples: they bring consumers to their search results pages and sell access to these consumers directly to advertisers; their platforms integrate all

TABLE 1 Evolution of online advertising formats

Advertising Format	2000	2001	2002	2003	2004	2005	2006	2007	2008*
	share of the total revenue coming from this format								
Display Related	78%	72%	60%	42%	39%	34%	32%	34%	33%
– Banners	48%	36%	29%	21%	19%	20%	22%	21%	21%
– Sponsorships	28%	26%	18%	10%	8%	5%	3%	3%	2%
– Rich Media	2%	2%	5%	8%	10%	8%	7%	8%	7%
– Slotting Fees	0%	8%	8%	3%	2%	1%	0%	0%	0%
– Digital Video	0%	0%	0%	0%	0%	0%	0%	2%	3%
Search	1%	4%	15%	35%	40%	41%	40%	41%	45%
Classifieds	7%	16%	15%	17%	18%	17%	18%	16%	14%
Lead Generation	4%	2%	1%	1%	2%	6%	8%	7%	7%
Email	3%	3%	4%	3%	1%	2%	2%	2%	2%
Interstitials	4%	3%	5%	2%	0%	0%	0%	0%	0%
Other	3%	0%	0%	0%	0%	0%	0%	0%	0%

*Figures are rounded

Ad Formats Definitions:
Banner is a space (usually rectangular) on a Web page that shows the advertiser's message.
Sponsorships represent custom content and/or experiences created for an advertiser which may or may not include ad elements. (e.g., reskinning a section of a web site with the advertiser's branding).
Rich Media refers to advertisements that incorporate animation, sound, and/or interactivity in any format.
Slotting fees the fee charged for premium ad placement and/or exclusivity.
Digital Video format includes commercials that appear in live, archived, and downloadable streaming content.
Search refers to advertising by paying Internet companies to list and/or link their company site domain name to a specific search word or phrase. It includes paid listings (text links appear at the top or side of search results for specific keywords), contextual search (text links appear in an article based on the context of the content, instead of a user-submitted keyword), and paid inclusion (guarantees that a marketer's URL is indexed by a search engine).
Classifieds: posting a product or service in an online listing for a fee.
Lead Generation are referrals to qualified purchase inquiries.
Email ads include banner ads, links or advertiser sponsorships that appear in commercial email communication.
Interstitials are ads displayed during a transition from one Web page to the next.

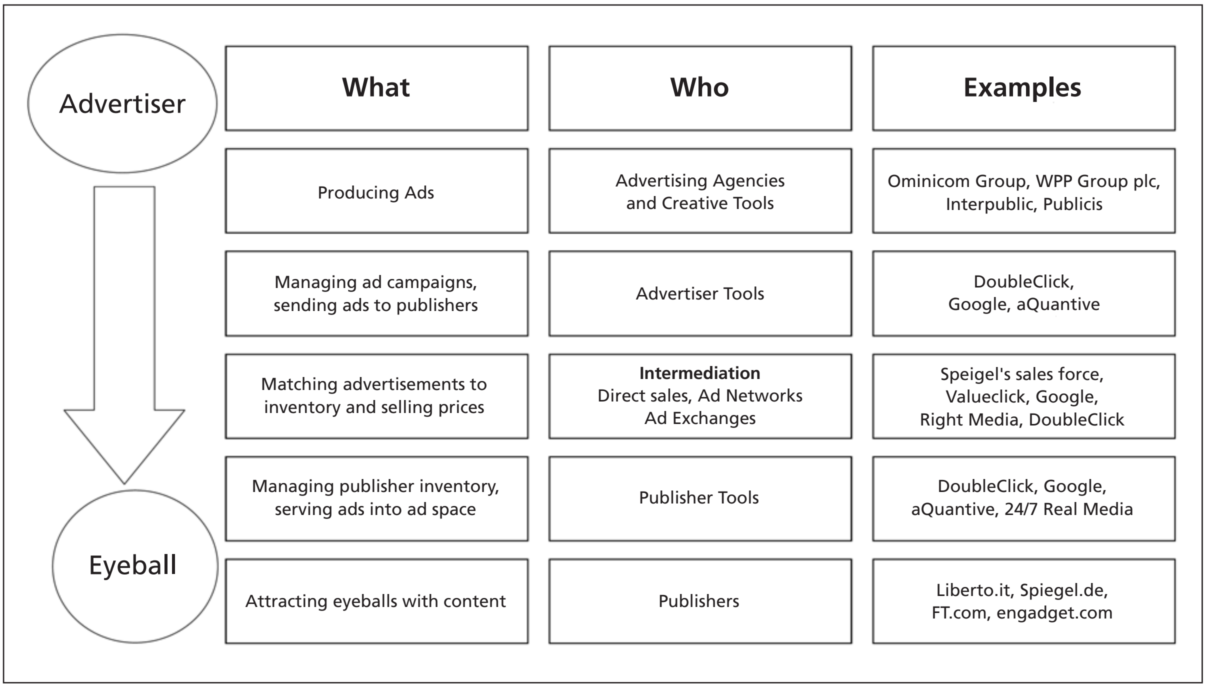
Source: Interactive Advertising Bureau Annual Reports and Press Releases, 2000-2008

the necessary technology for doing this. Many intermediaries are partly integrated. Publishers such as reuters.com bring consumers to their sites and have direct sales forces that sell advertising inventory directly to advertisers. But these publishers also typically rely on technology providers ("ad servers") that handle passing ads from the advertisers to the publishers' advertising spaces as well as advertising networks which aggregate online advertising inventory and sell it to advertisers. Finally, some publishers are highly specialized and contract most tasks out. That is true of blogs—even large ones—which rely on ad networks such as Google's Content Network to sell ads for them and to provide the relevant technology. Figure 1 shows the relationship between the various agents.

Almost all of the participants in intermediation between advertisers and consumers operate multi-sided platforms, sometimes working with agents for the advertiser or consumer. For example, media-buying firms work for advertisers and advertising agencies on the "buy side" and with publishers on the "sell side." This results in an industry of interlocking multi-sided platforms. Some of these platforms have more "sides" than just buy and sell. Facebook, for example, operates a software platform (Evans, 2009) that encourages developers to write applications that also enlist advertisers and consumers.

As with advertising generally, a key feature of online advertising is that consumers are paid with content and services to receive advertising messages and advertisers pay to send these messages. A fundamental question, not addressed here, is why this particular pricing and reward structure has held over long periods of time and across many different types of advertising. Among other things, the answer would help to illuminate the extent to which advertising is a method for reducing transactions costs between buyers and sellers or a source of imperfection that distorts decision making (Bagwell, 2002).

FIGURE 1 Relationship between various online advertising businesses



II. ONLINE VERSUS OFFLINE ADVERTISING

The fundamental differences between online and traditional advertising result from a combination of internet technologies and the nature of the web. The structure of online communications makes it easy for publishers and ad networks to learn considerably more about online users than has been possible with traditional media such as print, radio, and television. Online media or their ad networks typically know for certain from the internet technologies for linking people to sites that an individual is viewing their site. That is very different from a radio station, or a newspaper, which have limited ability to determine whether a particular individual is listening or reading. Online media can often learn valuable details about the individual that has signed on to the site. Each user has an IP (Internet protocol) address which typically identifies the location of the individual down to at least the zip code level in the United States. People who browse from home and from smaller companies typically have a unique IP address that remains the same over time. Using this address it is possible for online media and advertising networks to track other sites that users with that IP address have visited and to match up other details about the individual or household. (Some large companies change the IP addresses of individual users frequently so that the address cannot identify the user uniquely nor provide a precise geographic location. As a result online media cannot determine much about people who browse from these companies.)In addition, individual websites, such as wsj.com and myspace.com, may have detailed information on registered users which they can also use for advertising. Print, radio and television media generally do not know this level of information for individual users; cable systems with set-top boxes also have specific information on viewers (Lafayette, 2008) but do not have ready access to the browsing behavior of those individuals.

Traditional media usually have static scheduled content. Television and radio shows are broadcast at a particular time, newspapers are published daily, and magazines weekly or monthly. Advertisers and their intermediaries have no way of knowing whether an individual can hear or see their advertisements. Television viewers may leave the room and radio listeners may switch the channel when the ads come on. Readers may not necessarily look at particular newspaper and magazine pages with advertising. Consumers have much greater control over the content they view on websites. Advertisers and their intermediaries know with great confidence what content a consumer is viewing at a particular point in time. These facts have two implications. First, advertisements can be targeted to the particular view that is taking place. The platform can determine the time of day and location of the view and may also be able to determine various other characteristics of the viewer. Most on-line advertising inventory is selected in “almost” real time and customized for the particular viewer. The technology makes decisions on the ad to insert in a particular space so quickly that when you look at web page you cannot detect that it has been designed in less than the blink of an eye. Thus, advertisers have the ability to customize their advertising purchases in a way that is not cost-effective offline. Second, the advertising platform can often discern the context of why the viewer has come to the publisher’s web page. For example, search engines know the keywords a user requested, and publishers know the content of the page the user is looking at. Both may know recent search or browsing behavior. Ads can be customized based on this information.

These features can make online advertising a more efficient matchmaking vehicle for advertisers and viewers than offline advertising. Advertisers can target their messages to those consumers for whom the messages are most relevant and who are most likely to buy as a result of receiving this message. Viewers are more likely to receive messages that are relevant and valuable to them. An implication of this observation is that there are economic incentives for advertising and viewing to move online.

III. SUPPLY OF ADVERTISING SPACE AND ITS MARKET STRUCTURE

Any website that attracts viewers is a potential supplier of advertising inventory. Some websites choose to make money mainly in other ways: for example, e-commerce, gaming, adult sites, and a few others are largely free of advertising. But, as noted earlier, websites that account for a preponderance of page views among the top 100 sites earn most of their revenue from advertising. Table 2 shows the 20 largest advertising-supported sites in terms of page views, as of February 2008, and describes the content they use to attract eyeballs. Notable sites include Google, which primarily uses search results as well as user-uploaded video for its YouTube site; Yahoo!, which owns properties ranging from entertainment to automotive to email on which it presents ads; and Facebook, which operates a social networking site in which users see advertising on their own pages, the pages of their friends, and on other pages specifically designed for advertisers. The Table also reports 2008 advertising revenue for sites where this is publicly available—that ranges from a low of $130 million for Facebook to a high of $7.4 billion for Google.

The supply of advertising inventory is highly skewed. Thousands of web sites supply some advertising inventory in addition to the top one hundred. While comprehensive data on sites that supply advertising is not available, Figure 2 shows the overall distribution of page views, which is likely a good approximation of the distribution. As of February 2008, the top 10 sites accounted for 36 percent of page views, the top 300 accounted for 54 percent of page views, and top 10,000 websites attracted 67 percent of the total page views.

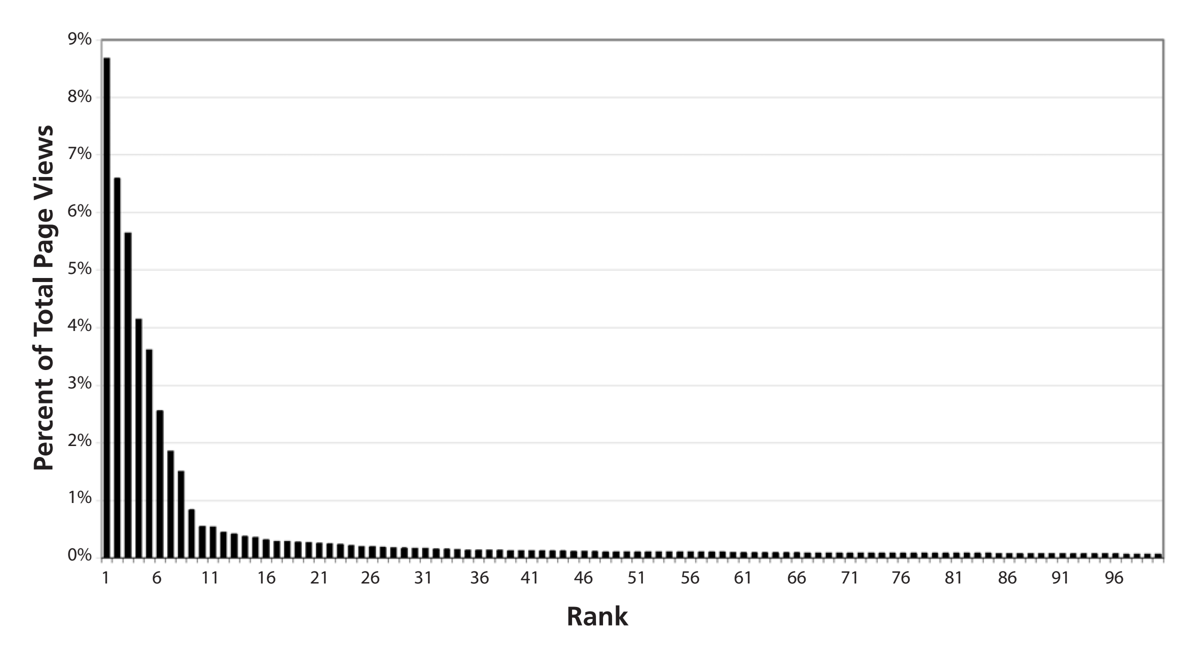
There are four large fully integrated suppliers of advertising inventory in the United States: AOL, Google, MSN, and Yahoo! Each of these sites is a "publisher" in the sense that it presents content which is used to attract viewers to their pages. Each also acts as a "distributor" of ad space by directly selling inventory to either advertisers or "brokers" that act as middlemen and match publishers with advertisers. Finally, these four sites supply most of their own technology. To be more specific, they operate the software and communication technology that takes copy from an advertiser and inserts it into space at the appropriate time for a viewer. In addition, MSN and Yahoo! are horizontally integrated into search; both of these entities have their search query tool incorporated in their portals.

TABLE 2 The top 20 web properties with ads and the content they use to attract viewers

Rank	Property	Content	2008 Internet Advertising Revenues ($M)
1	Fox Interactive Media	entertainment video, news, social networking, image hosting, games network	900
2	Yahoo! Sites	search results and various applications (news, email, weather forecast)	3430
3	Google Sites including YouTube	search results, email, maps, user-generated videos, blogs	7430
4	Microsoft Sites*	search results, email, entertainment videos, music, news	1970
5	AOL LLC	news, entertainment, email, search results, greetings	1360
6	FACEBOOK.COM	social networking site	130
7	eBay	online auction and shopping site for mostly used goods	2370
9	Comcast Corporation	TV listings, Free TV episodes, cable television services	
10	Viacom Digital	entertainment news, videos, music clips, TV listings, reviews	
11	Time Warner - (excl. AOL)	movies, TV schedule, videogames, cable television services,	
14	Amazon Sites	online shopping site, daily blog, customer reviews	
15	EA Online	strategy video games	
16	Ask Network	search results, public generated questions and answers, maps, news	480
17	Photobucket.com LLC	image and video hosting, sharing applications	
19	BEBO.COM	social media network, music, videos, applications	
21	Cox Enterprises Inc.	news, entertainment news, motor vehicles marketplace	
22	Disney Online	entertainment videos for kids, games, music library, fairy tale creating applications	270
24	United Online, Inc	floral product, internet services, social networking	
25	Glam Media	news and content on fashion, health and life style	
26	ESPN	sports news and videos, TV listings, sports results and rankings	

Notes:
1) Source: comScore Media Metrix data for February 2008.
2) Rankings are based on page views. Only web properties with ads are included. Some properties included well-known sites which are therefore not broken out separately; MySpace is included in the Fox properties and accounts for 98 percent of page view
3) Content information is determined through direct website reviews.
4) Although not all of Microsoft sites supply advertisements, e.g., MICROSOFT.COM, MSN.COM gains most of its revenue from displaying ads and accounts for 95% of all page views of Microsoft sites.

FIGURE 2 Distribution of page views for the top 100 websites



Source: comScore MediaMetrix February 2008.

Most of the other large suppliers are partly integrated. They usually rely on providers of publisher-serving technology for the software and communications system that receives copy from advertisers and then make decisions on where and when to insert that copy into inventory. They typically have their own sales forces that distribute their inventory to advertisers. But they also rely on advertising networks to sell space that they cannot or do not want to sell themselves (see the discussion below on price discrimination). Smaller websites ordinarily do not have enough volume to support a sales force. They often rely on an advertising network which may also provide the necessary technology. For example, a publisher can paste html code supplied by the advertising network into the part of the webpage the publisher wants to sell; that code will retrieve and insert an advertisement.

There are at least three main sources of supply of online advertising inventory, each of which results in different methods for selling advertising.

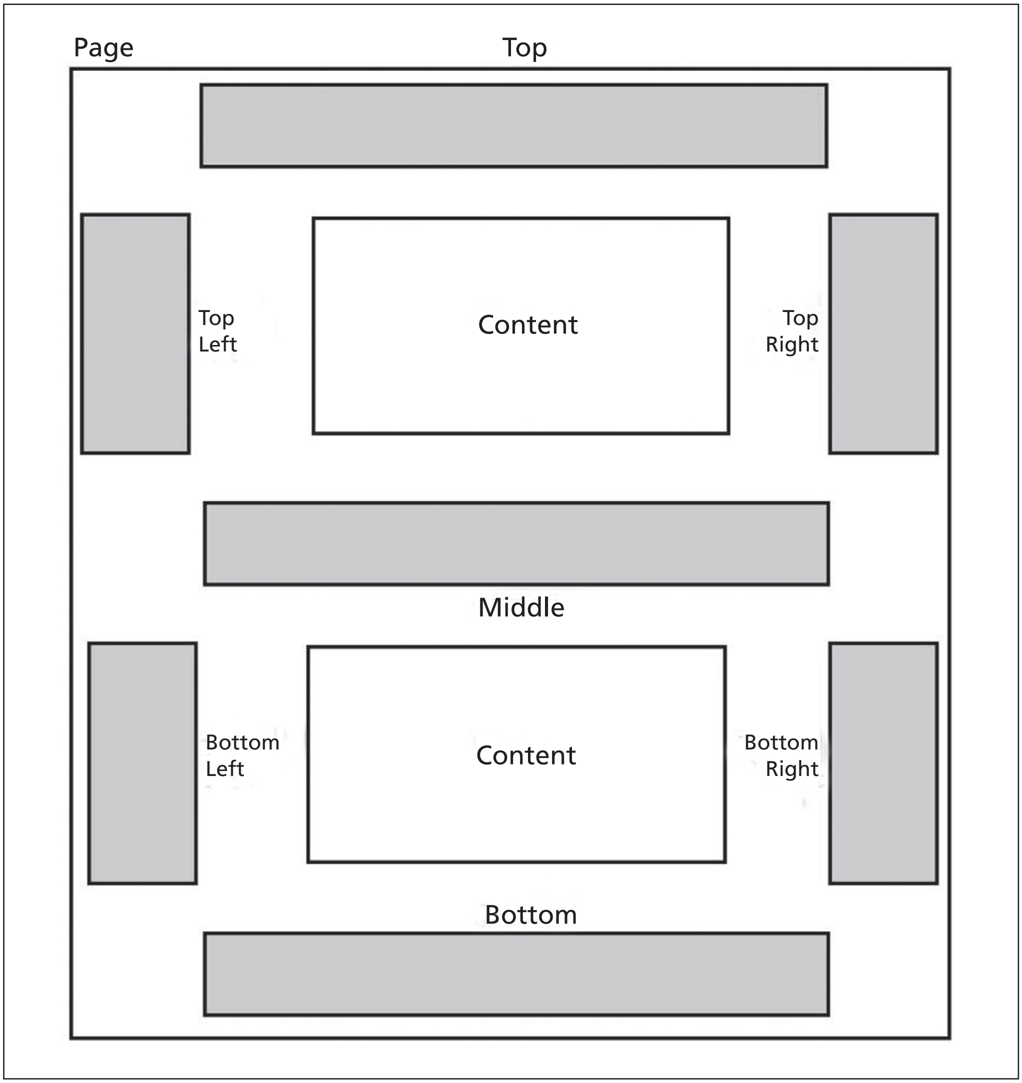
First, search engines generate search-results pages. Search engines need to enable users to find what they are looking for quickly and to move from the search results page to the desired webpage. In the late 1990s, search engines struggled to find the right balance between providing valuable search results to attract eyeballs and selling advertisers access to these eyeballs. Most ended up dividing the search results page into the "organic search results" that are based on the relevance of the web page to the keywords entered by the user; and the "paid search results" which are clearly demarcated text advertisements that also look like search results. Although organic search results are valuable to advertisers,[3] search engine providers do not charge advertisers for these listings. Instead, advertisers often hire "search engine optimization" companies, such as Performics, to increase their rankings in the search results. Each page typically has around ten slots available for a paid search advertisement. Advertisers bid on a cost-per-click for these slots and the search provider allocates the slots, roughly speaking, on the basis of expected revenue it will receive—that is, cost-per-click times expected clicks) (Evans, 2008; Varian, 2007). To maximize revenues from the scarce space, the search engines can use the cost-per-click bids and the expected number of clicks that the ads will receive to allocate the spots. Projecting the number of clicks and the effects of different allocation mechanisms is a difficult problem; the various search engine providers have solved that problem to varying degrees and their relative success in doing so explains in part differences in the revenue they receive for each search conducted (Evans, 2008).

Online media sites provide content that is broadly similar to what consumers get from traditional media. In fact, many traditional media companies have established web sites which use some material that is also provided through offline channels. Prominent examples include cnn.com, nytimes.com, mtv.com, and cosmopolitan.com. Other sites such as Yahoo! redeploy content from a variety of sources such as newswire services. Finally, some sites, such as youtube.com and drudgereport.com, provide content only online and often do so innovatively. With the exception of the video sites, these sites allocate portions of their pages to advertisements much like newspaper pages do. They generally sell the ads on a cost-per-mille basis—that is, based on how many eyeballs see their pages—through their own sales team and through advertising networks.

Figure 3 provides a sample layout: as with newspapers ads, different spots are perceived as more valuable because they are likely to receive more attention from consumers. Conversations with knowledgeable industry participants indicate that ads in the top half of the page garner a cost-per-mille of around $12.50 while ads in the bottom half of the page realize about half of that. Banners and skyscraper ads typically go for more than $12.50 cost-per-mille. The more desirable space tends to be sold directly, while the less desirable space tends to be sold through advertising networks. These ad networks pay between 15 and 45 percent of the cost-per- mille for a display ad on the top half of the page depending on the quality of the space.

[3] Search engines can be viewed as (at least) three-sided platforms based on third-party content obtained from the web, users looking for this content, and advertisers who seek access to these users. In principle, search engines could charge third-party content providers for the indexing and listing services made available but thus far have generally chosen not to. In March 2004 Yahoo launched a paid inclusion program that guaranteed listings on the Yahoo! search engine for commercial websites in return for payment. This scheme was not popular with website marketers or the public and was discontinued. Microsoft and AskJeeves charged for inclusion for a time but also discontinued this practice.

FIGURE 3 Sample layout of content and advertising on web page



People go to social networking sites to obtain updates on what their friends are doing, to update their own pages, and through these and other activities to communicate with their friends. These sites have allocated some space for advertising and have sold this inventory to advertisers. This space is sometime known as the "dead zone" because of the lack of attention that social networking users pay to it. The average cost-per-mille payments to social networking sites are often less than $0.50. The social networking sites, and on-line advertising businesses, are working on advertising methods that work more effectively with social networking communities. Google, for example, has developed a technology that makes it possible for advertisers to identify and target "influencers" in social networking communities to distribute messages (Helft, 2008). Whether social networking sites can achieve cost-per-mille payments for their viewers that are comparable to other types of on-line advertising remains to be seen. The promise of advertising on social networking websites is that word-of-mouth referrals are the primary influence on purchase decisions (BIGresearch, Simultaneous Media Usage Study (SIMM12), June 2008), and perhaps online communities can be used to create a source for such referrals.

The supply of advertising inventory is highly heterogeneous. It differs in size, the likelihood that consumers will pay attention to it, and the characteristics of the viewers. Not surprisingly, the price that advertisers pay per thousand views varies from a few cents to several hundred dollars in the case of high-income professionals. (This heterogeneity and price distribution is also true for traditional advertising.) Controlling for quality, however, suppliers of advertising engage in extensive "value-based pricing"—which economists call price discrimination.

The search-based advertising platforms, in principle, use second-price auctions to allocate slots and in this way seek to extract higher payments from those willing to pay more (Varian, 2007; see also Edelman et. al., 2007). Indeed, we might expect that the rents earned from the advertiser side would be bid away, at least partly, on the viewer side through competition among platforms (Rochet and Tirole, 2003; Armstrong, 2006), including through explicit subsidies to viewers to join the platform.[4] That has happened to a degree. Microsoft adopted some mechanisms for providing incentives to viewers in 2008; for instance, the Windows Live search site has a button labeled "Search a lot, Earn a lot" which describes incentives for searching that include free Xbox, Microsoft software, frequent flier miles, cash back and more (see <http://www.live.com/?form=MXCA00&kwid=019409a179d80c4ba7fe55c739b4f11c>, visited on November 28, 2008). Yahoo! has also entered into incentive arrangements for people to download toolbars that rely on its search engine. For example, Yahoo shares revenue with companies that get people to search with Yahoo and those companies in turn provide benefits—such as donations to charity—to individuals who download a toolbar that contains the Yahoo search feature (see <http://affiliates.freecause.com/index.cgi?action=about>, visited on November 28, 2008). Google has not followed suit. As of 2008, there does not appear to be robust price competition to persuade searchers to use alternative

[4] In general, multi-sided platforms can have profit-maximizing equilibrium price structures in which one or more sides have zero or negative prices (Rochet and Tirole, 2006). In the case of search-based advertising platforms the pricing on the advertiser side is fixed insofar as the platform adopts the standard auctions approach. These platforms compete, however, in a number of dimensions including the amount and quality of the search results provided to the users and the ease of use of the platform for users and advertisers. Such competition is similar to free television which competes through the attractiveness of the content rather than cash transfers to viewers. The search-based advertising platforms can dissipate the rents from the second-price auctions of advertising space spending on research and development and other expenditures that improve quality for users and advertisers. They can also dissipate it through distribution deals such as those discussed in the text.

engines available to them. A different kind of competition for attention does take place. Google, Microsoft and Yahoo! compete to become the default search engine on the browser, on various tool bars that are distributed, and in publisher web sites. Although consumers can readily change these default settings the search engine providers presumably pay for these because enough consumers stick with the default. These deals provide incentives to entities that can provide search engine defaults for their users to compete harder to these users by providing greater quality.

Online publishers engage in price discrimination as well. This practice is similar to other business-to-business transactions which are often characterized by individual negotiations, and the online practices are similar to the offline practices. In both cases, heterogeneous supply is being matched up with heterogeneous demand through bilateral exchanges and with limited public information. Media companies can take space they do not choose to sell directly and make it available to intermediaries that sell it indirectly. Larger online publishers typically have contracts with several advertising networks and have technology that selects the highest price for inventory in more or less real time. Online advertising networks are discussed in more detail below.

IV. DEMAND FOR ONLINE ADVERTISING SPACE

Businesses spend money on advertising because it helps them sell products to consumers. Economists have developed a variety of models concerning the role of advertising, but most of that work focuses on how advertising might affect demand. There is little empirical work that assesses the extent to which advertising spending reduces transactions costs by providing valuable information (including quality signals) to consumers and matching buyers and sellers; provides disinformation that harms consumers; or is designed to alter preferences. In addition, few studies have examined the demand for advertising from the vantage-point of how the advertisers decide how much to spend or how to allocate that spending across different forms of advertising (Silk et al., 2002). Existing work together with anecdotal information suggest that advertisers—and their agents—determine an overall advertising budget, allocate that budget among different methods (such as brand advertising on national television) for achieving the objectives of an advertising campaign, and select various alternative advertising outlets for spending their dollars. Traditional methods have included placing ads on different media such as television, radio, newspapers, magazines, billboards, and directories and engaging in sales as mail and telemarketing. Each method draws upon many suppliers of advertising inventory, which vary in their coverage of the population as well as the characteristics of those covered.

Advertisers typically develop "campaigns," which utilize a mixture of these methods and suppliers, to achieve objectives which may range from increasing sales of an existing product, to introducing a new one, to affecting the image of the company or brand. They base their decisions on the level and allocation of their budgets on formal or informal analyses of the rate of return on investment (Duboff, 2007). For these ad campaigns, the different advertising methods can be substitutes to the extent they provide alternative ways of delivering messages to an audience and complements to the extent they can reinforce each other ("The Medium Is The Massages," 2002). Berndt et al. (2002) find that 57 percent of the 28 pairs of the cross-elasticities they estimated indicated the methods were, on net, substitutes and the remainder were complements--although typically weak ones.

The emergence of online advertising has provided another type of media for advertisers. The online methods of advertising, however, usually have close counterparts offline. A GM Corvette ad on cnn.com is not visually that different from a GM Corvette ad in *USA Today* and GM will pay for both ads based on the number of viewers. A Yahoo! ad for pizza restaurants on the Upper West Side of Manhattan is not that different from an ad for a pizza restaurant in the Manhattan "Yellow Pages" telephone directory. Hardcopy mail and e-mail solicitations for vacuum cleaners are similar, and both entail payment based on results. Of course, potentially important differences may arise in the value received by the advertiser from the comparable online and offline message. Online, the GM ad viewer can click through to a website; someone from out-of-town can find the pizza restaurant on Yahoo; and the e-mail solicitation for vacuum cleaners can take a buyer to a website to purchase the vacuum and have it shipped.

Different methods of online advertising are potentially substitutes or complements both with each other and with other forms of advertising. For example, if an advertiser wants to reach a large number of individuals to introduce a new product, it can buy a banner ad on the Yahoo! home page which has 48 million visitors a day (based on comScore Media Metrix report for February 2008) or it can buy a couple of 30-second television spots on Fox TV's *American Idol* which had more than 20 million viewers for its early 2008 episodes. A typical advertiser will place display ads on multiple web sites as well as several television shows to reach a large audience. However, as these methods are all designed to reach large numbers of people, they are probably substitutes at the margin. The advertiser might complement this campaign by buying keywords on one or more of the major search engines so that consumers that see the product can search for it online, learn more about it, and possibly purchase it. Of course, investigating the fine structure of demand requires more than such ruminations. The extent to which different methods of offline and online advertising substitute or complement for each other depends on the nature of the particular campaign, the objectives the advertiser seeks to pursue, the aggregation across advertisers of many possibly varying demand relationships, and other factors. The net degree of substitution or complementarities between different advertising methods is an open empirical question.

Advertisers typically hire firms to design and execute advertising campaigns. Although these firms often have relationships through a conglomerate, such as WPP, they have become more specialized over time (Berndt et al., 2008). Increasingly one firm does the creative work and plans the campaign while another firm engages in buying and placing media. While some firms have arisen that specialize in online advertising, most creative work and media buying is performed by advertising firms that manage both offline and online work for advertisers. Within the online part of the business, advertisers or their agents also purchase various technologies for distributing online advertisements to suppliers of advertising inventory and measuring the success of online campaigns.

V. INTERMEDIATION BETWEEN THE ADVERTISING BUY AND SELL SIDES

The advertising industry has to solve a massive matching problem between businesses and consumers. A large number of advertisers want to deliver multiple messages to a large number of consumers. Indeed, advertising agencies were formed in the mid-nineteenth century to deal with the coordination of supply and demand among businesses that wanted to advertise outside their locality and the daily and weekly newspapers (Pope, 1983).

The online advertising industry has developed a variety of technologies and business methods for solving this matching problem. The most innovative and best-known involve "search-based advertising," in which advertisers and consumers are matched based on the "keywords" that people enter into search engines. Consumers are attracted to the search engine because they are interested in content on the web. Advertisers use these keywords as proxies for the likelihood that consumers would be interested receiving a message that might lead to a sale. Auction methods are then used to allocate the advertising inventory to the businesses that are willing to pay most for most clicks, which in turn depends on the amount they are willing to pay and the number of clicks their ads will receive (Evans, 2008).

Other innovations are occurring in the buying and selling of display advertising on publisher websites. As of 2008, the preponderance of display advertising (measured by revenues) was bought and sold the old-fashioned way. Large online publishers have salespeople who sell their inventory to media buyers for large advertisers. That method of selling advertising space is only viable when there are enough people viewing that space—and thus enough expected revenue—to warrant the cost of salespeople calling on individual advertisers. Computer-based methods for matching the supply of advertising inventory and the demand for advertising inventory have made it possible for smaller sites to sell their inventory profitably. These methods also enable larger publishers to sell inventory that their salespeople have not sold. A number of advertising networks have arisen which broker multilateral exchanges between publishers and advertisers. These advertising networks enter into agreements with publishers to sell available advertising and with advertisers to deliver viewers with specified characteristics in return for a fee. Some of these networks provide behaviorally targeted ads discussed below although many place ads based on crude demographic information. Google has developed a computerized solution that has proved economic for many small web sites such as blogs. Google's Content Network supplies advertising inventory from "hundreds of thousands" of web sites that have joined its network (according to <https://adwords.google.com/select/afc.html?sourceid=awo&subid=en-us-et- awhp_related&hl=en_US>, visited December 7, 2008). In what is known as "contextual advertising," Google's advertising network auctions keywords that appear on the web page of participating publishers and inserts ads from participating advertisers based on the appearance of those keywords and possibly some other characteristics of the website. The publisher pastes html code into its webpage to receive and display the ad, while the advertiser typically uses an auction and advertising campaign management tool that is bundled into the software package it uses for search campaigns. (Sears, 2005). Many larger websites also use Google's contextual advertising for some of their less desirable space.

It appears likely that online advertising provides two potentially significant economic efficiencies. First, a promising conjecture is that online advertising allows the economy to reduce the amount of resources devoted to creating content for aggregating and sorting potential buyers. Society may not need to invest as much in magazines, newspapers, and other media whose main purpose is aggregating the right eyeballs for advertisers. Although consumers value that content they have not had to pay for its cost of production. Second, online advertising likely increases the accuracy of the match between the buyer and the seller. The seller has greater ability to target consumers that are likely to buy and the consumer is more likely to receive useful messages and less likely to receive time-consuming but irrelevant messages.

In both cases one can argue that there are some losses that also need to be considered. The news media gathers and reports news through professional journalists in some cases scattered through the world including war zones. One can argue that the news media provides an important public service in a democratic society and that its value exceeds what individuals or advertisers may pay for it. Moreover, as with any technological change some people gain and others lose. Those who strongly prefer the touch and feel of newspapers will lose if not enough people are willing to support the costs of these newspapers. In addition, online advertising may provide more efficient matching and delivery of ads but that begs the longstanding question of whether advertising is providing people with valuable information that helps them make better buying decisions, or whether it is getting people to buy based on deceptive information or by persuading them to do things that they will soon regret.

VI. ISSUES IN INDUSTRY EVOLUTION

Online advertising is one of those "gales of creative destruction" that will reshape several industries and radically change traditional ways of delivering advertising messages from sellers to prospective buyers.

VII. BEHAVIORAL TARGETING AND DATA ANALYTICS

As noted earlier it is possible for online entities to gather data on what people have done online including what they have searched for, what web sites they have browsed, and perhaps even what they have purchased online. Those data together with other information on these people can be used to target advertisements to people based on their behavior.

For example, an advertising intermediary could help an automobile insurer target individuals who probably have good risk profiles, who may be buying an expensive new car, and who are therefore likely to in the market for automobile insurance. The intermediary could infer that the individual may have bought new expensive car from the fact that the individual has been browsing particular websites that people go to when they are going to purchase a luxury automobile. The intermediary might also be able to infer from online behavior that the individual falls into a low-risk insurance category. It could infer from a user's IP address and browsing behavior that the user is probably a woman (from browsing behavior) who lives in a well-off suburb (from the IP address), in a region with low accident and theft rates (from the IP address), and is over the age of 25 (from browsing behavior). The intermediary might identify potential sales targets by just using simple screening methods such as these or may use more sophisticated statistical methods to predict the sort of the prospects the insurance company is looking for. Either way the advertising intermediary can use these methods to determine whether a particular individual who is browsing a website at a particular moment has exhibited the web browsing behaviors and personal characteristics that make that person a good target for an ad. It can make that decision almost instantaneously and then insert an ad into advertising space on the page that the target is viewing. The advertiser would typically pay a premium over standard online advertising rates for views by these targeted individuals because the likelihood that they will "convert" the view by this "qualified prospect" into a sale and thus their expected profits is higher.

Although behavioral targeting is an area of intense innovation, as of 2009 only a small portion of the advertising revenue earned by publishers results from selling these sorts of behaviorally targeted advertisements.[5] Two factors limit the current deployment of this seemingly efficient method of advertising.

First, since behavioral targeting narrows the group of people that see an advertisement the likelihood that these individuals will ultimately purchase the product has to be high enough to offset the reduction in the number of people that view that advertisement. On average only about 1 out of 400 viewers click on a given ad (Marketing Sherpa, 2008) and only a fraction of those viewers purchase the product. Unless behavioral targeting is sufficiently precise advertisers may prefer to reach a larger group of individuals all else equal. Reaching 10,000 people of whom 1/1000 (or 10) will ultimately purchase a product is better than reaching 2,000 people of whom 1/500 (4) will ultimately purchase the product. There is a tradeoff between precision and reach. In this example, all else equal, a behavioral screen that targeted a fifth of the potential population would have to identify people that were more than five times as likely to buy to be better than targeting everyone. Precision is limited by the amount and quality of data that are available. (In addition, behavioral targeting is not relevant for brand advertising that is generally aimed at a broad audience to influence their views on a company or a product rather than to make a direct sale.)

[5] Behaviorally targeted advertising is mainly used by advertising networks. Advertisements placed by advertising networks account for a small portion of the advertisements on publisher web sites and only a portion of the advertisements placed by advertising networks rely on behavioral targeting. Based on discussions with industry participants the fraction of revenue for advertisements placed by advertising networks is small and behavioral targeting smaller still.

Second, the advertising platform that implements a behavioral targeting campaign must have access to a large enough universe of viewers to find enough candidates to make the campaign worthwhile to the advertiser. There are fixed costs of designing and executing advertising campaigns; the advertiser needs to make enough sales to recover these costs and make a return. Suppose that it costs $10,000 to design a campaign and the expected incremental profit from each sale that results from this campaign is $10. Then the campaign would have to generate an expected 1000 conversions to recover the fixed costs. If only 1 out of 1000 consumers that are exposed to an ad is converted to a sale, the campaign would need to reach at least 1,000,000 individuals for it to break even. Consider a behavioral targeting campaign that converts 1 out of 500 consumers by targeting the 20 percent of the potential universe of people that are the most likely buyers. To yield 1000 conversions that campaign would need to reach 2,500,000 people. The total sample needed depends generally on the conversion rate and the targeting rates which depend on each other. Many advertising networks are not large enough to engage in highly refined behavioral targeting while some of the largest advertising platforms have not yet deployed highly refined targeting.

Several developments may increase the use of behavioral targeting methods. With improvements in predictive techniques and the availability of data on viewers behavioral targeting would become more precise and advertisers would increasingly prefer targeting to reach. If advertising networks that use behavioral targeting methods increase their scale, or if larger advertising platforms increasingly deploy behavioral targeting and data collection methods, they will obtain enough viewers to make behavioral campaigns economically efficient. Privacy concerns discussed below could limit the development of behavioral targeting. Consumers may resist having advertising platforms collect detailed information about their browsing behavior and government regulations may limit the ability of advertising intermediaries to collect these data.

VIII. ELECTRONIC EXCHANGES

Online publishers rely on advertising networks to sell inventory that they have not succeeded in selling directly, and also as a substitute for direct selling in some cases. Some advertising networks enter into deals with online publications to sell some of their inventory to advertisers. The advertising network then sells this space to advertisers. Some networks might sell the advertiser the ability to deliver their ads into a variety of similarly situated advertising inventory such as financial publications without guaranteeing which ones; others might sell the advertiser the ability to deliver ads into selected publications; and still others might sell the advertiser the ability to target particular types of users. Some advertising networks focus on broad publications while others concentrate on particular niches. Aside from the ability to serve ads electronically, these exchanges follow roughly the same business model as the nineteenth- century advertising agencies that brokered advertising space (Pope, 1983).

Several firms have developed or are proposing to develop electronic exchanges (for one example, see the discussion the DoubleClick Advertising Exchange at <http://www.doubleclick.com/products/advertisingexchange/index.aspx>, visited on November 29, 2008). These exchanges connect web-based publishers (or their agents) that would make their advertising inventory available and advertisers (or their agents) that would want to purchase this inventory. Advertising networks that have an excess supply or demand of advertising inventory could also participate. The advertising inventory would be auctioned off in real time and the exchange then handles the delivery of the advertising from the advertiser to the publisher. No data on the size of these exchanges is available, but based on discussions with individuals in the online advertising business, as of the end of 2008 it does not appear that the existing exchanges have achieved a critical mass of liquidity that would enable them to sustain themselves. Indeed, few such business-to-business exchanges have achieved critical mass in any area of business—most have died (Evans, 2009). It remains to be seen whether the advertising exchanges will ultimately face a similar fate. (Harris, 2002 discusses the role of liquidity for an exchange and Evans and Schmalensee, 2009 discuss the role of critical mass for multi-sided platforms.)

IX. CREATIVE DESTRUCTION AND THE MIGRATION OF OFFLINE TO ONLINE

Traditional advertising sustains a complex ecosystem of businesses. A wide range of media entities earn significant portions—sometimes all—of their revenues from the sale of advertising inventory. These include newspapers, magazines, free television, free radio, billboards, and yellow pages. In turn, these businesses support a variety of content generation businesses, including television production companies and musicians. Diverse other businesses work with advertisers including advertising agencies, media buyers, and audience measurement firms. Every business in this ecosystem felt a breeze as on-line advertising arrived in the mid-1990s, and then felt a stiff wind by the early 2000s as the online advertising industry came together.

Online advertising methods pose a serious threat to traditional methods for several reasons. First, they increase the efficiency of matching buyers and sellers and delivering advertising messages to the buyers. In the long run one would expect that this will reduce the economic importance of traditional intermediaries such as advertising agencies, media buyers and sellers, and direct sales forces.

Second, they increase the supply of advertising inventory significantly. By providing a method for earning revenue from attracting viewers, online advertising attracts the entry of content providers that supply advertising inventory. During the 2000s that has included of user-generated video sites such as YouTube, social networking sites such as Facebook, and blogs such as HuntingtonPost. Americans viewed more than 458 billion web pages in 2007, each of which could have carried advertising and many of which did. The aggregate number of page views increased at an average rate of 21 percent between February 2004 and February 2008 (according to ComScore Media Metrix). There are low financial barriers to the formation of these sites as a result of web technologies. This increased supply of advertising inventory puts downward pressure on advertising rates, promises to reduce the returns that traditional media can get from advertising, and therefore potentially reduces the quality-adjusted supply of content by traditional media. (The increase in the quality of advertising could increase or decrease the demand by advertisers which could therefore have a partly offsetting effect on rates.)

Third, online advertising increases the supply of online content which provides a substitute for traditional content. The potential returns from online advertising encourage entities such as Yahoo! and MSN, as well as traditional media, to present various kinds of content online that consumers used to consume mainly offline. In addition, of course, viewers are moving from offline to online media because, as with the move from radio to television, they simply like the content better along certain dimension, which can include the ease, flexibility, and interactive dimensions of access. Thus far the evidence shows that as a result of online content consumers have substituted away from radio and newspaper content but not television content. The US Census reports that persons above 12 years of age spend 32% more time watching that cable and satellite television than they did in 2001. The usage of radio and traditional newspapers has declined by 3% and 15%, respectively.[6]

These trends are likely to accelerate as more content is available either online or through other technologies that enable the sellers of advertising inventory to employ online methods. Some traditional media may make a fairly smooth transition to the online world. The television industry could change its delivery method so that programming is delivered primarily through internet protocols—what is known as IPTV—and implement online technologies to serve ads. Verizon's FiOS Television, which is based on internet technologies, had 2.5 million Internet subscribers and 1.9 million television customers in the United States as of December 31, 2008.[7] The production and programming sides of the business could operate much like they do today although many of the business methods, and assorted institutions, for selling and delivering advertising would change. The same is true for radio which is already being delivered over the internet. One cannot consume internet radio on the morning drive to work at the moment but that may change once there is more extensive wireless internet coverage and cars become equipped with internet-enabled electronics.

The forces of creative destruction have hit the newspaper industry the hardest and the earliest. The industry is in free fall as a result of the self-reinforcing migration of readers and advertisers

---

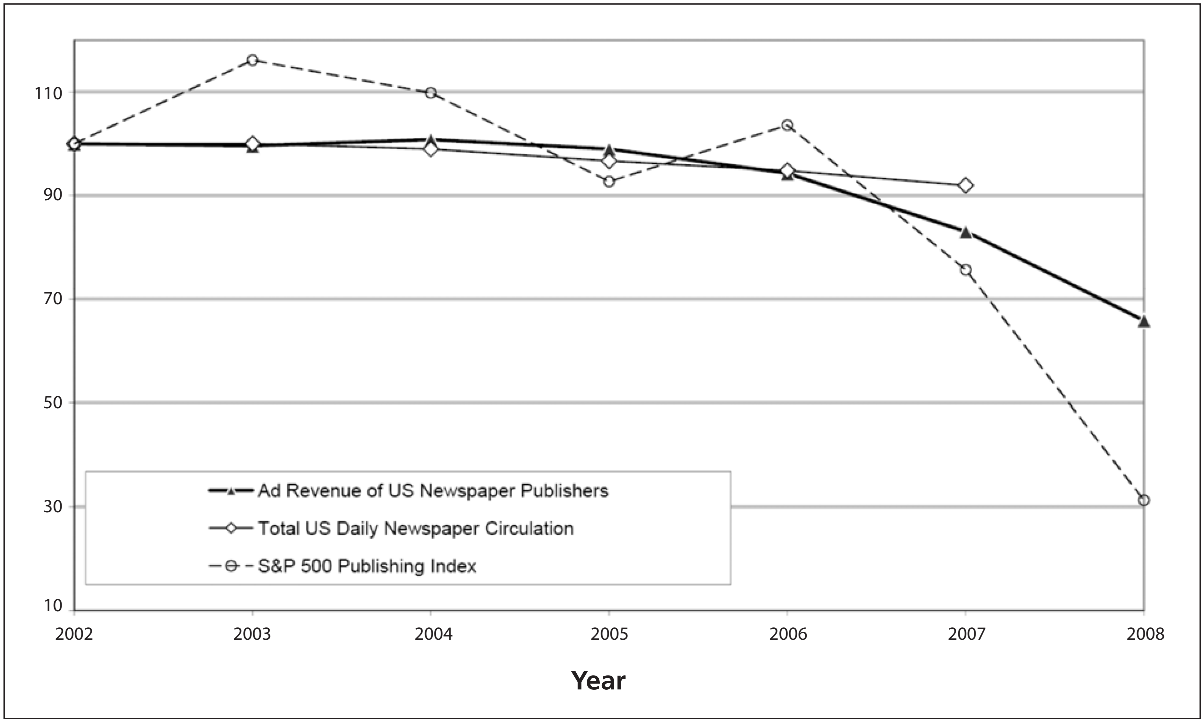
[6] US Census Bureau, Table 1089: Media Usage and Consumer Spending: 2001 to 2011, http://www.census.gov/compendia/statab/tables/09s1089.pdf.

[7] "Verizon FiOS-Fact Sheet," Verizon News Center (available at http://newscenter.verizon.com/kit/fios-symmetrical-internet-service/all-about-fios.html, last visited on March 24, 2009).

online. Figure 4 shows the circulation for all U.S. daily newspapers, their advertising revenues, and the market capitalizations of the Standard & Poor's 500 Publishing Index—all indexed to 100 in 2002. There has been a precipitous drop in circulation and advertising revenue and the S&P 500 indicates that the market expects dramatic reductions in the future profit stream of the newspaper industry. As mentioned earlier several major daily newspapers in the United States have closed their doors. Although many newspapers have developed web versions that attract significant number of viewers, their increased online advertising revenues have only partly compensated for their loss of traditional advertising revenues. As radio and television move to internet-enabled platforms they could find the same decline in advertising revenues with the same impact on the economics of providing traditional content.

The industrial structure of the online advertising industry could evolve in a variety of ways. One possibility is that the industry will have a highly concentrated set of intermediaries at its center,

FIGURE 4 Newspaper Industry Performance Measures



Ad revenue and publishing index are adjusted for inflation, all quantities indexed to 2002 base year. In 2002 ad revenue of all US newspapers equaled $44 billion, and the total US daily newspaper circulation was 50.7 million.
S&P 500 Publisher Index is composed of the prices of the common stocks of the following companies: Gannett Co., Inc., The McGraw-Hill Companies, Inc., Meredith Corporation, The New York Times Company, and Washington Post Company. It is weighted by market cap.
In 2002, the total market cap of these companies was $44.1 billion.
Source: Advertising Age, Newspaper Association of America, Washington Post Company Annual Reports.

with many content providers around this core. This result would be akin to the U.S. stock exchange system which has 5 major exchanges that account for 92 percent of transaction volume.[8] Another possibility is that the industry will have many intermediaries at its center. Some of the intermediaries will focus on mass advertising while others will focus on niches. The ultimate structure depends on the relative importance of several factors: the strength of indirect network effects and scale economies on one side, and the possible benefits of specialization of knowledge in certain areas.

X. THE PRIVACY DILEMMA

Participants in the online advertising industry collect and store a great deal of information about people who use the internet. Search-engine providers capture every search a user does with their search engines and the websites they visited. They store these data in a way that enable them to identify the individual IP address that generated the data for several months. Many participants in the online advertising industry also insert "tracking cookies" into the computers of users with whom they have had contact. These cookies enable the provider that inserted them to track the web sites that an individual has visited. These providers may also capture these data and store them along with the IP address that identifies the user. "Web beacons" are code on web pages that determine that an individual has opened a page and captures the IP address of that individual. These beacons can track the activity of the individual on a site. The traditional advertising industry also collects data on people and uses that for targeted mailings and telephone calls. The online advertising industry collects vastly more data.

The providers that collect these data use them for various purposes. In some cases they employ them to conduct research to improve their products. Google's privacy page, for example, says that it uses them to conduct research on common spelling mistakes and to assess the effectiveness of the ad rankings. Past browsing behavior is also critical to the implementation of the behavioral targeting strategies described above. These data can be used to infer general characteristics of individuals such as their gender and interest in sports as well as their present buying interests such as whether they are planning a vacation.

These detailed data on browsing are valuable to online advertising industry providers. They enable these providers to provide higher quality prospects to advertisers and to therefore charge more for the advertising inventory they supply. They also possibly provide more valuable advertisements to users who in turn will be more likely to visit web sites that present advertisements that are more tailored to the needs of the user. Given that a web site is going to display an advertisement consumers might prefer that the advertisement be more relevant than not.

---

[8] Data as reported by Arcavision. Note that the 5 exchanges are associated with either Nasdaq or NYSE Euronext. The shares of each of the five exchanges are: NASDAQ-29.9%, NASDAQ TRF-19.8%,NYSE Arca 18.1%, NYSE 17% and NYSE TRF 7%.

The collection of these data has proved controversial, however, and raises some delicate public policy issues. Consumer privacy advocates and regulators have criticized the search-engine providers for capturing and storing data (Dye 2009). In response, these providers have agreed to reduce the length of time for which they store data with individual identifiers. Google, for example, has reduced this period from two years to nine months.[9] The collection of data, as well as behavioral targeting, has attracted lawsuits and legislative inquiries. NebuAd and Phorm are good examples. The two behavioral targeting firms enlisted Internet Service Providers into their advertising network and used browsing data from these Internet Service Providers to target advertisements. They quickly became the subject of legislative inquiries in the United States and the United Kingdom and many of the Internet Service Providers withdrew as a result of the controversy.[10]

The use of personal data for targeted advertising raises several public policy issues. If people had ownership over information about themselves, and there was a competitive market for it, they could decide whether to sell it to an online advertising business. In making this decision, people could take into account the value of receiving possibly more relevant ads, as well as any other compensation they might get. They would also take into account any costs from the possible leakage of their private information.

To a degree, consumers have some control over their private information. A user can choose not to use websites that insert "cookies" that collect data on her machine (or choose to delete such "cookies" on a regular basis). Web browsers have increasingly provided mechanisms for consumers to control the retention of information on their browsing history and manage their cookies. Reviews of the recently released browsers in the market, IE 8.0 and Google Chrome 2.0, emphasize new features like private browsing and search suggestions. [11] To the extent the consumers exercise these choices they put competitive pressure on online advertisers to account for the value that consumers place on data about themselves. Consumers can also avoid websites and web services scan and store personal content. Some users may avoid Google Gmail because it scans and records the content of emails while others may value the targeted ads that result from Google's ability to examine the content of the emails.

Nevertheless, three potential problems arise that could warrant the consideration of government intervention into the treatment of privacy by the online advertising industry.

First, there is the usual imperfect information problem that is often used to justify consumer protection efforts (Stiglitz and Walsh, 2002). Consumers may not know that information is being collected and stored. It appears likely that few consumers knew that Google stored each user's

[9] "Google to dump user data earlier," BBC News (September 9, 2008), available at <http://news.bbc.co.uk/2/hi/technology/7605801.stm>

[10] See Paul (2008).

[11] <http://www.microsoft.com/windows/internet-explorer/features/overview.aspx?tabid=1&catid=1>, <http://www.google.com/chrome/intl/en/features.html#>

search history, including details sufficient to identify the IP address of the user's computer, for two years before this was the subject of news stories (for example, "Google's growth raises privacy concerns," 2008). Even with the publication it remains unclear how much consumers know about the extent to which various online advertising related businesses collect information on them.

Second, consumers may agree (either tacitly or explicitly) to provide private information without anticipating that this information would be sold to other vendors who might combine it with other information about them. For example, a consumer may feel differently about providing search data associated with her IP address to a search engine provider if she knew the search engine provider would sell the data to another vendor that had figured out a way to associate her IP with her personal information including name, address, and telephone number.

Third, competition among advertising platforms may not necessarily result in the optimal provision of privacy. When there is perfect information competition can usually be relied on to yield the optimal provision of quality to the consumer. Online advertising intermediaries are multi-sided platforms that compete simultaneously for advertisers and viewers. Whether this competition results in the optimal provision of privacy, and the extent to which it would do so in a highly concentrated market, would need to be investigated carefully (see Rochet and Tirole, 2006) for an analysis of similar issues in credit cards).

In principle, of course, consumers could learn about the collection and use of their information. But this incurs costs. Likewise, the online advertising businesses could increase efforts to engage in transparent contracts with consumers. But this incurs costs, too. The critical public policy question is how property rights—including those enforced through regulation—over private data should be assigned.

Public policymakers in the United States and the European Community are grappling with these issues (Federal Trade Commission, 2006, 2007; European Commission, 2008). Too stringent regulations could harm consumers. After all, the online advertising industry benefits consumers in several ways. It increases the likelihood that they will receive relevant ads and decreases the likelihood that they will waste time on irrelevant ads. Moreover, it promises to reduce the costs of advertising to businesses, and some or all of these costs would be passed on to consumers in the form of lower prices. On the other side, overly lenient regulations could also harm consumers. Consumers could incur the costs of having private information disclosed and potentially misused, and incur the costs of reducing their use of the web because of concerns over privacy. Regardless of whether it their private information is disclosed consumers may not like receiving ads that reflect too much knowledge about them even if it is only a software program on a remote server that has that knowledge.

Resolving the privacy dilemma is important for innovation in behavioral targeting. Innovators will benefit from knowing what information they can collect and how they can use it without risking lawsuits, being pilloried in the press, and being hauled in front of Congress. Consumers will benefit from balancing the benefits of receiving relevant ads against the cost of losing valuable privacy.

XI. APPENDIX

Statistics for Online Advertising of the Top 100 Ranked Websites

Of the top 100 ranked website: percent of total page viewed	 47.84%
Websites with ads of the top 100 ranked websites: percent of total page viewed	56 41.32%
Of those with most revenue from online ads: percent of all page views for the top 100	26 77.14%

Rank	Top 100 Ranked Properties	Percent of total pages viewed	Supply Ads?	Most Revenue from Ads
1	Fox Interactive Media	8.67%	1	1
2	Yahoo! Sites	6.58%	1	1
3	Google Sites	5.64%	1	1
4	Microsoft Sites	4.14%	1	1
5	AOL LLC	3.61%	1	1
6	FACEBOOK.COM	2.55%	1	1
7	eBay	1.85%	1	1
8	craigslist, inc.	1.50%	0	0
9	Comcast Corporation	0.83%	1	0
10	Viacom Digital	0.54%	1	1
11	Time Warner - Excluding AOL	0.53%	1	1
12	Intuit	0.44%	0	0
13	AT&T, Inc.	0.41%	0	0
14	Amazon Sites	0.36%	1	0
15	EA Online	0.35%	1	0
16	Ask Network	0.31%	1	1
17	Photobucket.com LLC	0.28%	1	1
18	Bank of America	0.28%	0	0
19	BEBO.COM	0.27%	1	1
20	Verizon Communications Corporation	0.26%	0	0
21	Cox Enterprises Inc.	0.25%	1	0
22	Disney Online	0.24%	1	1
23	TaxACT	0.23%	0	0
24	United Online, Inc	0.21%	1	0
25	Glam Media	0.19%	1	1
26	ESPN	0.19%	1	1
27	Wal-Mart	0.18%	1	0
28	Pearson Education	0.17%	0	0
29	Wikipedia Sites	0.16%	0	0
30	GAIAONLINE.COM	0.16%	1	0
31	JPMorgan Chase Property	0.16%	0	0
32	Weather Channel, The	0.15%	1	1
33	Weatherbug Property	0.15%	1	1
34	Gorilla Nation	0.14%	1	1
35	Gannett Sites	0.13%	1	0
36	Earthlink	0.13%	1	0
37	SINGLESNET.COM	0.13%	1	1
38	Expedia Inc	0.12%	1	0
39	New York Times Digital	0.12%	1	1
40	Target Corporation	0.12%	0	0
41	Apple Inc.	0.12%	0	0
42	Wells Fargo	0.12%	1	0
43	JCPenney Sites	0.12%	0	0
44	Community Connect, Inc.	0.11%	1	0
45	The Generations Network	0.11%	1	0
46	E.W. Scripps	0.11%	0	0
47	People Media Sites	0.11%	1	0
48	CNET Networks	0.10%	1	0
49	WildTangent Network	0.10%	1	0
50	CBS Corporation	0.10%	1	1

Rank	Top 100 Ranked Properties	Percent of total pages viewed	Supply Ads?	Most Revenue from
51	PROJECTPLAYLIST.COM	0.10%	1	0
52	Sprint Nextel Corporation	0.10%	0	0
53	WorldNow - ABC Owned Sites	0.10%	0	0
54	CareerBuilder LLC	0.10%	1	1
55	PLENTYOFFISH.COM	0.10%	1	0
56	ADP.COM	0.10%	0	0
57	MYHOTCOMMENTS.COM	0.10%	1	0
58	Citigroup	0.10%	0	0
59	Match.com Sites	0.09%	1	0
60	H&R Block	0.09%	0	0
61	FRIENDSTER.COM	0.09%	1	1
62	Monster Worldwide	0.09%	1	0
63	HI5.COM	0.08%	1	1
64	Limitedbrands	0.08%	0	0
65	BLACKBOARD.COM	0.08%	0	0
66	NETFLIX.COM	0.08%	0	0
67	DEVIANTART.COM	0.08%	1	0
68	Midasplayer.com Ltd.	0.08%	1	0
69	Capital One	0.08%	0	0
70	TD Bank Financial Group	0.08%	0	0
71	Ford Motor Company	0.08%	0	0
72	Liberty Media Holding Corporation	0.08%	1	0
73	Macy's Inc.	0.08%	0	0
74	FUN Technologies	0.08%	0	0
75	Hearst Corporation	0.08%	0	0
76	Ticketmaster	0.08%	1	0
77	ONEMANGA.COM	0.08%	1	0
78	Wachovia Corp	0.08%	0	0
79	Dell	0.07%	0	0
80	IRS.GOV	0.07%	0	0
81	AmericanGreetings Property	0.07%	1	1
82	HOTFREELAYOUTS.COM	0.07%	1	0
83	Clear Channel Online	0.07%	0	0
84	eHarmony	0.07%	1	0
85	Washington Mutual	0.07%	0	0
86	PBS	0.07%	1	1
87	REALPAGE.COM	0.07%	0	0
88	ATHENAHEALTH.COM	0.07%	0	0
89	Groupe PPR	0.07%	1	0
90	Cablevision	0.07%	0	0
91	Sears Sites	0.07%	0	0
92	Avon Products Inc.	0.07%	0	0
93	UPS Sites	0.07%	0	0
94	Move Network	0.07%	0	0
95	iVillage.com: The Womens Network	0.07%	0	0
96	Best Buy Sites	0.07%	0	0
97	ED.GOV	0.06%	0	0
98	Yellowpages.com Network	0.06%	1	1
99	Staples.com Sites	0.06%	0	0
100	Travelport	0.06%	0	0

Source:comScore Media Metrix February 2008
Each website was reviewed between November 15th and December 5th to determine whether it presented advertising on the home page or immediate branches of the home page.

CHAPTER 9

Web Economy, Two-Sided Markets and Competition Policy

David S. Evans

ABSTRACT

The web economy has grown rapidly in the last decade. Online businesses have several key features that are important for understanding the pro-competitive and anti-competitive strategies they may engage in. The two-sided markets literature helps elucidate many of these strategies. It also provides guidance for the antitrust analysis of market definition and exclusionary practices for web-based businesses.

Ten years ago a tweet was something a bird did. We generally did not poke our friends. When we sent an email about buying a car we would have jumped out of our chairs if an advertisement for BMW all of a sudden popped up on our screens. And our mobile phones did not have application stores. Change has occurred rapidly following the development of the commercial internet but it has a long way to go. It takes years for entrepreneurs to come up with ideas, for businesses to start, and for industries to evolve and sort themselves out. It will take a decade or two, perhaps, for things to settle down. Competition policy will find itself increasingly grappling with mergers, exclusionary practices, and collusion as the web economy becomes more prominent, as it matures, and as it goes through a period of significant turmoil.

A decade ago we would have finished the phrase "*two-sided*" with "coin" and not markets and we would have thought that a person who engaged in multihoming had a place in the country. Like the web economy the study of two-sided platforms has grown rapidly since its birth at the turn of the century. The literature has flourished with many theoretical papers and much empirical research. Most major competition authorities around the world are using the two-sided market approach in a broad range of cases from credit cards to shopping malls.

Competition policy matters involving web-based businesses will provide fertile ground for using two-sided analysis. In this paper I want to use the lens of two-sided markets to describe some key features of the analysis of market definition, market power, and exclusionary practices particularly as they relate to entry.

I. A TOUR OF THE WEB ECONOMY

The web economy is constantly evolving. Today businesses fit into one or more of the following categories. (1) E-commerce includes massive shopping malls like eBay, Amazon, and Baidu. It also includes many retailers that have set up shop online such as walmart.com. (2) There is online media which includes everything from portals like MSN to online video like YouTube, to newspapers like The Guardian.com, to all of you that have blogs. (3) Social networking is the new kid on the block on the internet. Many of you have profiles on Facebook or Bebo or some kind of site like that. (4) Online gaming has become enormous. It ranges from social networking games like Farmville, which you can find on Facebook, to Xbox Live.

Many of these web-based businesses make money from attracting eyeballs and selling access to those eyeballs to advertisers. This is where on-line advertising comes in. Many web sites run advertising sales themselves just like traditional newspapers and magazines have. That includes Yahoo and reuters.com. But then there are many businesses on the web that act as advertising intermediaries. They operate networks of advertisers and media properties and they pool inventory. Advertising is important for another reason on the web. The e-commerce and the media properties are advertisers themselves. What could be better for online advertising to drive clicks to them ? AOL has become a very big player.

We now take a brief bus tour of some of the more interesting properties on the web.

Google, of course, has to be the first stop. We hear so much about this ten-year-old company that we need to put it in some perspective. Google makes virtually all of its money from selling

TABLE 1 A Deep Dive into the Web Economy

eCommerce	Online Publishing	Social Networking	Online Advertising
Auction sites eBay	***Portals*** Yahoo, MSN	***Mass Personal*** Facebook, Myspace	***Integrated*** Yahoo, MSN, Google AdWords
Online malls eBay, Amazon	***Print Publishers*** Nytimes.com, CNN.com	***Professional*** LinkedIn	***Intermediaries*** Google Adsense, advertiser.com, tacoda.com
eTailers Tigerdirect.com, walmart.com	***Video publishers*** Youtube.com, ifilm.com	***Niche personal*** Asmallworld.com	
	Blogs Huntingtonpost.com		

Note: Web categories overlap considerably. Social networking uses self-generate content to attract eyeballs that can be sold to advertisers. Search-results pages are also content that attracts advertisers. Some businesses are "mash ups" of different categories; for example YouTube relies on self-generated content and social networking to attract eyeballs.

text-based ads on search-results pages and a bit more with contextual ads on its publisher network. It is not making much money from many of the other businesses that it has gotten into and it has largely abandoned its efforts to make money from selling traditional ads for radio and television. It is an important player, nevertheless, for three key reasons. First, it's the dominant player in helping people find things on the web and that is terribly important. Second, it gets billions of dollars of revenue a year from search and contextual ads and this money helps it fund grander ambitions. Third, it does not like other firms controlling parts of the web because that gets in the way, or can get in the way, of it selling more advertising.

The iPhone is the next stop. What makes the iPhone so important is that it is a platform for developing applications. There are more than a 100,000 applications that people can download for their phones. More of these are being written every day and most are web-based applications. These applications are helping to transform other industries—below we will see the example of Square which may disrupt payments.

Facebook is a six-year-old company that is run by a twenty-five-year-old. It has taken over the social networking space but it is actually a late entrant in that segment. There was a flurry of social networking sites that started in the late 1990s such as Six Degrees of Separation. Then Friendster was the king of the segment in the early 2000s. MySpace toppled Friendster and then Facebook has leap-frogged MySpace. Like YouTube, Twitter and other sites that attract eyeballs, Facebook really has not figured out how to make huge dollars from the huge traffic that it generates around the world. The idea is advertising. The problem is that when you go visit your friends you do not necessarily want to have a stranger trying to sell you male enhancement drugs. If you are an advertiser like Procter and Gamble you may not want your ads appearing next to pictures of drunk and half-naked people.

How do these web celebrities get along in the neighborhood? Google doesn't like Facebook because it cannot get its spiders (which pull content for its search engine) onto this large and growing part of the web. Facebook likes the iPhone because it is a great device for people to go visit their friends. Google and Apple used to get along so well that Google had a couple of representatives on the Apple board. But then Google could not stand having Apple control a large part of the mobile space for inserting ads. As a result, Google unfriended Apple and they are now engaged in the mobile platform wars.

II. KEY FEATURES OF THE WEB ECONOMY

The web economy has many interesting economic features. I would like to highlight several that are important from a two-sided perspective.

A. Critical Mass

We have seen a lot of entry of platforms over the last 15 years. For every one that succeeds, and that you have heard about, far many more have failed. There were more than 40 video sites that secured enough viewers to be counted around the time that YouTube began back in 2005. Almost all of them are gone now. More generally, most platforms fail to take off. The growth of these platforms is driven by network effects and they face a critical mass problem. The problem is analogous to an exchange which can survive only if it gets enough liquidity. Much of the work that web entrepreneurs like the founders do when they start is trying to figure out ways to get enough customers on board the platform to take off. YouTube as an example of tackling the critical mass problem. It had to figure out how to get enough people to upload videos and how to get enough people to view those videos and how to get both of those groups onboard the platform in enough numbers to ignite the platform and to get rapid growth. Businesses that get enough liquidity (such as Hulu and YouTube) ignite while those that don't (such Revver, another video sharing site) impode. It is as if there is an invisible wall. Once the entrepreneur pushes the platform through that wall the platform can take off.

We have known for a long time in a two-sided literature that exclusivities are a way to solve the chicken and egg problem. The reverse is true as well. One way for an incumbent firm to prevent a new firm from taking off is to make it hard for it to get enough critical mass by entering into exclusives with enough of at least one major customer group.

B. Free

Many of the web-based platforms discussed above are free to at least one group of customers. That is a well studied and documented phenomena in the two-sided literature. The platform generates value by getting one or more customer groups together. It can be profitable to charge one group of customers nothing just to get them on board so that the platform can charge another group of customers for access to them. We often see this offline. There are many free newspapers, for example. We often see « free » online because the marginal cost of serving another user is zero. Thus, you do not have to pay for your Facebook page. That social network makes money by selling your eyeball to advertisers and selling complementary goods like virtual gifts.

There is a tendency in competition policy cases to ignore the customers that are getting things for free. That happens for two reasons. One is that analysts tend to equate the business with the money side. We have seen this with Windows. An important group of customers for Microsoft are software developers that use the Windows APIs. Microsoft has chosen to provide most of its services to these customers for free. But those customers get a lot of value from Microsoft and should be considered in any welfare calculation. Another reason that analysts ignore the free side is that the traditional methods of market definition focus attention on a single group of customers even though the members of the two groups have welfare that is inextricably intertwined.

C. Invisible Engines

If you looked for the heart and soul of the web business what would you find? It is not the server farm in South Dakota with all the lights flashing. It is the software. If you decided to start a new web business like Pandora or Skype or Twitter, you would mainly spend your time writing software code. All of these businesses on the web are based on thousands of lines of code. This software can be locked down so no one can get access to it. It can also be opened up so that others can use the features of it ; in that case it becomes a software platform that can support the development of complementary applications. Many of the web-based businesses have started software platforms by opening up their code. You might wonder how Firefox managed to cut Internet Explorer down to size. Much of it had to do with encouraging developers to write applications that increased the value of the Firefox browser. Facebook has done the same thing and there is an active developer community on Facebook with more than 500,000 applications written so far. Farmville is an example of a Facebook application. Google Maps has become such a powerful product because Google made the APIs available so developers could build applications that integrate mapping.

The software platform model is transforming the web. Almost every major web property has followed that strategy. It is propelling rapid growth and innovation.

From a competition policy perspective these applications that have been built on top of these platforms cut two ways. For one, they are the source of great value. The developers benefit from the platform directly because it makes it possible for them to engage in innovation, to write applications for your iPhone or Facebook or for Google for example, at a low cost. The consumers of the applications also benefit. But these applications also pose a possible barrier to entry. It is the old chicken and egg story. It takes a lot of effort to compete with the incumbent platforms that already have both sides onboard. As a result, if you want to compete with the iPhone, you have to cope with the fact that the iPhone has a 100,000 and counting applications.

D. Mashups and Morphing

We see a lot of mash-ups on the web. That means creating new services by combining things. Square is an example. That is a new payment system that was created by Jack Dorsey, who is one of the founders of Twitter. He has a software application that works with the iPhone. A small merchant can add a square attachment to your iPhone and can accept cards (after singing up for a processing agreement). Consumers who swipe their cards in that square device enter their emails and become part of the network. This new platform provides an alternative payment system.

It is also relatively easy for web businesses to morph rapidly in ways that few might have expected. You might think that LinkedIn is like Facebook. It is not. LinkedIn is a job board and recruiting tool. It makes money basically by acting as a recruiting tool, selling job postings and so forth.

Mash-ups and morphing are important for analyzing market definition and market power. A few years ago TomTom, which was a leading supplier of handheld navigational devices, bought TeleAtlas which was one of the few major suppliers of maps used for navigation. The Commission approved the merger but it was controversial. Google has just completely disrupted that business since then by mashing up its Android phones and Google Maps. The Android phone can be an incredibly powerful navigational tool that people can use in their cars and

TABLE 2 History of Social Networking Leaders

Social network	Birth	Unique visitors Nov 2009	Status
Six Degrees of Separation	1997	0	Dead
Friendster	2002	2 million	Very popular in the Philippines
MySpace	2003	62 million	Gasping for air
Facebook	2004	103 million	In heat
???????	????	??????????	Under the radar screen or yet to be invented

pretty much everywhere at zero marginal cost. The stock price of TomTom and Garmin which is another navigational device maker plummeted last October after the Android phone came out with this navigational device.

E. It's Only Just Begun

There is a tendency to think that we are at the end of history. The latest new thing is the last thing. That is what people thought when Friendster created a successful social network. And that is what they thought about MySpace when it killed Friendster. Is Facebook really the final thing ? At some point we will be at the end of history for some of these categories. Some company will nail it and will dominate for a long period of time until something completely new comes along. But it takes a long time and only in retrospect to know when that happens.

Even then, the pace of innovation in this area should give one pause on how secure any dominant position is on the web. No one predicted in the early 2000s that Google would be playing rope-a-dope with Microsoft or that Microsoft would be complaining about someone else leveraging their monopoly power.

III. COMPETITION POLICY FOR THE WEB ECONOMY

There is no reason that competition policy cannot deal with all the issues that are going to emerge in the web economy over the next few years. The web is hardly alone in being more complicated and different than Adam Smith's pin factory. Moreover, many of the two-sided issues that arise for the web occur in traditional industries, such as physical exchanges, payments and shopping malls.

If there is one place where the analysis could go wrong it is market definition. Done correctly market definition ought to be a tool for understanding competitive constraints and helping to evaluate unilateral and coordinated effects. The problem arises when mechanical methods are used to draw hard market boundaries and when the perspective of this market provides a distorted view of the competition that is actually taking place in the real world.

There is a growing consensus among economists that the current tools for market definition face numerous problems especially when there is product differentiation which there almost always is. The hypothetical monopolist test is quite difficult to implement reliably. The results are highly dependent upon whatever assumption the economist is making about the shape of the demand schedule and the sequence of the products considered among other things. All those problems become much more difficult when the markets are two-sided. Market definition, for example, does not deal very well with the complementary products that characterize two-sided markets. Many of the problems that critical loss has in one-sided markets become an order of magnitude harder in two-sided markets.

For the web economy competition policy will do better to rely on methods that are less formulaic and based more on qualitative research into the nature of competition in the business ecosystem.

As with any new kind of business the web businesses will provide creative ways of monopolizing unlawfully that we have not thought of. There will also be pro-competitive explanations that we have not imagined for practices that look suspicious. I mentioned critical mass earlier. Firms use many methods to achieve critical mass and obtain platform ignition. Firms can use these same sorts of methods to prevent their rivals from getting critical mass. It may not take much to prevent a rival from achieving ignition. That is something for competition authorities to be concerned about.

Google will likely become the testing ground for the next decade of antitrust analysis of the web economy. The economics in this company is really pretty simple. It makes money from advertising. That means it wants spaces to put ads and it wants eyeballs to look at those ads. The more space and the more eyeballs they have the better. My suspicion is that Google enters other parts of the web primarily to ensure that nothing comes between them, the advertising space, and the eyeballs looking at that space. We are seeing this now in mobile. Advertising and eyeballs are going to mobile devices. Google has started the Android operating system and launched it own line of phones to help ensure access to mobile advertising inventory. As Google keeps entering other parts of the web ecosystem, we are going to return to the an debate. Is Google a dominant firm, leveraging its way into other businesses. Or is this a double marginalization story where a dominant firm in one market can make consumers better off by either making a related market competitive or extending its own dominance into that market. Thus, is it an anti-competitive story or an efficiency one.

The web economy is intellectually interesting, the issues are complicated, and there is much room for debating whether practices are good or bad. All of this will make the practice of competition policy most intriguing and perhaps profitable, for good or not, for many years to come.

CHAPTER

10 Antitrust Issues Raised by Global Internet Economy

David S. Evans

ABSTRACT

Web-based businesses are increasingly the subject of antitrust concerns. Plaintiffs in the United States have sued eBay for tying its online payments service to its transaction service. Multiple jurisdictions in the European Community have claimed that Apple has violated the competition laws by limiting the ability of its music player to play music from competing music stores and limiting the ability of competing music players to play music purchased from its music stores. During 2007, although the U.S. Federal Trade Commission decided not to block Google's acquisition of DoubleClick after a lengthy investigation, it expressed its intent to "closely watch these markets" involved in online advertising. The web economy poses two major challenges to competition authorities. The law and economics for analyzing the multi-sided platforms that dominate the internet sector is not well developed. At the same time the web-economy is evolving very rapidly and in ways that are sure to result in antitrust complaints and investigations. Competition authorities and courts will need to exercise great care in balancing the protection of consumers from anticompetitive behavior against causing harm from interfering in complex businesses that are both rapidly moving and not fully understood.

Web-based businesses are increasingly the subject of antitrust concerns. Plaintiffs in the United States have sued eBay for tying its online payments service to its transaction service.[1] Multiple jurisdictions in the European Community have claimed that Apple has violated the competition laws by limiting the ability of its music player to play music from competing music stores and limiting the ability of competing music players to play music purchased from its music stores.[2] During

---

[†] This Essay was previously published in the *Northwestern University Law Review Colloquy* on May 5, 2008, as David S. Evans, *Antitrust Issues Raised by the Emerging Global Internet Economy,* 102 NW. U. L. REV. COLLOQUY 285 (2008), http://www.law.northwestern.edu/lawreview/ colloquy/2008/13/.

[1] *See, e.g.,* Complaint, Malone v. eBay, Inc., No. 07-01882-JF (N.D. Cal. Apr. 4, 2007).

[2] Consumer Council of Norway, Complaint against iTunes Music Store (Jan. 25, 2006), *available at* http://forbrukerportalen.no/filearchive/Complaint%20against%20iTunes%20Music%20Store.pdf; Helena Spongenberg, *Apple's iTunes Risks EU Backlash After Norway Ruling,* EU OBSERVER, Jan. 25, 2007, http://euobserver.com/871/23334; *Apple: European Consumer Organisations Join Forces in Legal Dispute over iTunes Music Store,* VERBRAUCHERZENTRALE, Jan. 22, 2007, http://www.vzbv.de/go/dokumente/571/8/36/index.html; *Free-DRM Music by iTunes, but EC Starts Official Investigation,* EDRI.ORG, Apr. 12, 2007, http://www.edri.org/edrigram/number5.7/ituned-free-drm; *Increased Competition Pushes iTunes Towards Interoperability,* FORBRUKEROMBUDET, Feb. 11, 2008, http://www. forbrukerombudet.no/index.gan?id=11038653.

2007, although the U.S. Federal Trade Commission decided not to block Google's acquisition of DoubleClick after a lengthy investigation, it expressed its intent to "closely watch these markets" involved in online advertising.[3]

Of course, competition policymakers have not just discovered the Web. In 1998, shortly after the start of the commercial Internet three years earlier, the U.S. Department of Justice and various states filed an antitrust case against Microsoft for engaging in various practices related to Web browsers.[4] The European Commission started an investigation of Microsoft's practices related to media players that stream music over the Internet in 2001.[5] However, the Microsoft cases mainly involved the use of the company's market power in personal computers to influence competition in Web-based markets that threatened it. The matters involving Apple, Google, and eBay concern market power in Web-based products and services themselves.

The Internet economy is likely to raise antitrust concerns—and possible demands for regulation—for years to come. Gargantuan global firms have emerged, which will likely attract scrutiny by competition authorities and by policymakers concerned with competition issues. The companies mentioned above, for example, have shares in putative antitrust markets that rival those held by Microsoft.[6] Apple has more than a 70% share of paid music downloads in the European Union,[7] Google has more than an 80% share of search queries in Europe,[8] and eBay has more than a 90% share of auction site page views in France, Germany, Italy, Spain, and the UK.[9] Competition authorities and private parties can challenge the practices of these leading firms under the antitrust laws of most jurisdictions. Such challenges are especially likely under European Community law and decisional practice, which impose special obligations and significant scrutiny on firms that have market shares as low as 40%.[10]

---

[3] Fed. Trade Comm'n, Statement of Federal Trade Commission Concerning Google/DoubleClick (Dec. 20, 2007), *available* at http://www.ftc.gov/os/caselist/0710170/071220statement.pdf.

[4] The D.C. Circuit Court of Appeals upheld in 2001 portions of a lower court ruling in 2000 that found Microsoft guilty of violating section 2 of the Sherman Act. See United States v. Microsoft Corp., 253 F.3d 34, 45 (D.C. Cir. 2001) (citing United States v. Microsoft Corp., 87 F. Supp. 2d 30 (D.D.C. 2000)).

[5] In 2007, the European Court of First Instance upheld the Commission's 2004 decision that Microsoft had violated Article 82 EC by tying its media player software to its dominant operating system. See Case T-201/04, Microsoft Corp. v. Comm'n, 2007 WL 2693858 (Sept. 17, 2007).

[6] These are not necessarily relevant antitrust markets, but they are ones that competition authorities could plausibly adopt.

[7] *See* Anna Jenkinson & Don Jeffrey, *Apple, Record Companies Restrict Music Sales, EU Says (Update6),* Bloomberg.Com, Apr. 3, 2007, http://www.bloomberg.com/apps/news?pid=20601085&sid=aiMdS7KddG14&refer=europe.

[8] comScore, MyMetrix qSearch 2.0 Key Measures Report (Dec. 2007). comScore is a "global Internet information provider" that gathers data on Internet usage trends. comScore, Who We Are, http://www.comscore.com/about/default.asp (last visited Aug. 12, 2008). comScore's data analyses are based on its panel of over two million users. Id. In recruiting its panelists, comScore attempts to ensure that "[a]ll demographic segments of the online population are represented in the comScore Global Network, with large samples of participants in each segment." comScore, comScore Methodology, http://www.comscore.com/method/method.asp (last visited Aug. 12, 2008).

[9] comScore, MyMetrix Key Measures Report (Dec. 2007).

Moreover, many Web-based firms have complex business models and arrangements. Separating the merely complicated from the nefarious will take courts and competition authorities time to sort out. This chapter describes the economics and technology behind the Web-based economy and how these features will influence Internet competition policy in the years to come.

Part I provides a birds-eye view of the Web-based economy. Although this sector is evolving quickly, its contours are beginning to take shape and we can be reasonably confident that several globally dominant firms will play significant roles. Part II describes the economics of the Web-based economy. The key businesses are what economists call "multi-sided platforms" that serve several distinct but interdependent customer groups. Google, for example, serves people who are searching the Web, advertisers who want to reach these users, and application developers who are using Google's software to develop complementary products. The leading multi- sided platforms for the Web are often built on "software platform" technologies that make portions of their code available to software developers who write applications that benefit users of the software platform. Part III considers the competition that arises in the Web-based economy. The appearance of dominant firms in key sectors will ensure ongoing scrutiny, and the nature of the economics and technology of these businesses will result in ongoing disputes over their practices.

The Web economy poses two major challenges to competition authorities. Currently, the law and economics for analyzing the multi-sided plat forms that dominate the Internet sector are not well developed. At the same time, the Web economy is evolving very rapidly and in ways that are sure to result in antitrust complaints and investigations. Competition authorities and courts will need to exercise great care in balancing the protection of consumers from anticompetitive behavior against causing harm from interfering in complex businesses that are both rapidly moving and not fully understood.

I. AN OVERVIEW OF THE WEB-BASED ECONOMY

The Internet refers to computer networks that are linked through wired and wireless connections and that interoperate through standard communication protocols. This global communication system provides access to various software-based services. The most important set of services on the Internet is the World Wide Web, which consists of digital media linked through hyperlinks

[10] *See* Case T-219/99, British Airways plc v. Comm'n, 2003 E.C.R. II-5917, ¶¶ 211, 223–25 (find ing British Airways dominant in the context of Article 82 with a share of 39.7%, notwithstanding evidence that its share had declined from 46% over a seven-year period).

The finding in *British Airways* relied heavily on the fact that the rest of the market was very fragmented. See id. ¶¶ 211–25. This was the first time that a firm with a market share below forty percent was found to be dominant. Subsequently, in *Wanadoo Interactive*, the Commission concluded that Wanadoo did hold a dominant position, though it only had a market share of 39%. Case COMP/38.233, Wanadoo Interactive, Commission Decision of July 16, 2003, ¶ 227. The Commission reached this finding based on the size and strength of Wanadoo's main competitors, who all had markets shares in between 6.5% and 16%. Id. The issue of thresholds for finding dominance was not examined further by the Court of First Instance in Case T-339/04, *France Télécom SA v. Comm'n,* 2007 E.C.R. II-00521.

and uniform resource locators (URLs).[11] For simplicity, this chapter refers to the Internet as the physical communication system and to the Web as all products and services that rely on this communication system.

The key innovations behind the Internet started occurring in the early 1960s, and the key innovations behind the Web came in 1989.[12] The commercial Web began roughly in 1995 with the introduction of browsers that made Web navigation easy for regular computer users; this therefore encouraged the formation of businesses geared to a mass audience. There was much prognostication in the late 1990s on how the Web would evolve and enormous optimism that it would lead to quick fortunes as network effects—more users make a site more valuable which leads to more users—propelled early entrants to monopolies.[13] Those hopes were seemingly dashed in 2001, when the market capitalizations of most Web-based firms plummeted and vast numbers of these firms vanished.[14] It became apparent that many of these firms had not developed business models that allowed them to make money from the visitors who came to their sites. In the aftermath of the burst bubble, a robust Web-based economy has emerged that is creating completely new services, from social networking to behaviorally targeted advertising, and also disrupting many traditional businesses from media to telephony.[15] As we will see below, advertising revenue for delivering traffic has become the major driver for a significant portion of the Web economy.

A large portion of the population in industrialized countries touches the Web-based economy daily. In 2007, 55% of the population in the European Union[16] and 71% of the population in the United States had access to the Internet.[17] In December 2007, 78% of all Internet users in the

---

[11] Other Internet services include online gaming, Voice-over-IP, email, instant messaging, file sharing, and other communication services. Most of these services are bundled into Web sites.

[12] *See* David C. Mowery & Timothy Simcoe, *Is the Internet a U.S. Invention?—An Economic and Technological History of Computer Networking*, 31 RES. POL'Y 1369, 1371–76 (2002).

[13] *See*, e.g., CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 13 (1999) ("When the value of a product to one user depends on how many other users there are, economists say that this product exhibits network externalities, or network effects . . . . Technologies subject to strong network effects tend to exhibit long lead times followed by explosive growth. The pattern results from positive feedback: as the installed base of users grows, more and more users find adoption worthwhile.").

[14] *See*, e.g., Suzanne McGee, *New Economy Sours on Venture Capitalists,* WALL ST. J., Jan. 2, 2001, at R6; *Easy.com Easy.gone*, ECONOMIST, June 16, 2001, Special Section: A Survey of the New Rich, at 6; *Living in Freefall*, ECONOMIST, Nov. 18, 2000, at 117–18.

[15] *See Is 'Web 2.0' Another Bubble?*, WALL ST. J. ONLINE, Dec. 27, 2006, http://online.wsj.com/article/SB116679843912957776.html.

[16] Internet World Stats, European Union Internet Usage Stats and Population Statistics, http://www.internetworldstats.com/stats9.htm (last visited Apr. 19, 2008). Population penetration is defined as the number of Internet users divided by the total population of the region. An Internet user is defined as "anyone currently in the capacity to use the Internet," namely, that he has access to an Internet connection point and can use the technology. Internet World Stats does not adjust the figures to exclude children, illiterate people, or any other segment of the population.

[17] Internet World Stats, North America Internet Usage Stats and Population Statistics, http://www.internetworldstats.com/stats14.htm#north, (last visited Apr. 19, 2008).

United States and 69% worldwide used webmail;[18] 39% of all Internet users in the U.S. and 47% worldwide used instant messaging.[19] According to a 2007 survey of Americans who frequently use the Internet, 84% used a search engine.[20] The same survey found that the average respondent spent thirty-three hours per week on Web products and services in 2007 compared with sixteen hours viewing television.[21]

Web-based products and services are consumed primarily through the personal computer, which is the primary Internet-connected device in most countries.[22] However, it is widely expected that most mobile phones around the world will soon be connected to the Internet.[23] This increased portability will increase the amount of time people can access the Web because people usually have their mobile phones with them all the time. It will also increase the use of Web-based products and services in lesser developed countries because mobile phones are less expensive than personal computers and more widely held.[24]

One must be modest in speculating on the future evolution of the Web. The Internet and the Web are very new technologies by historical standards. One could not have reliably forecasted the development of electricity at a similar point during its development. The recent dot-com bust teaches how wrong smart and financially motivated people can be about the business prospects of a new technology. Nevertheless, five features of the Web-based economy appear to have emerged

---

[18] *See* comScore Media Metrix, Web-Based Email Usage for December 2007, http://www.marketingcharts.com/wp/wp-content/uploads/2008/02/comscore-yahoo-microsoft-google-email-web-based-us-worldwide-december-2007.jpg (last visited Aug. 12, 2008).

[19] *See* comScore Media Metrix, Internet Messenger Usage for December 2007, http://www.marketingcharts.com/wp/wp-content/uploads/2008/02/comscore-yahoo-microsoft-google-instant-messenger-us-worldwide-december-2007.jpg (last visited Aug. 12, 2008).

[20] IDC, U.S. Online Consumer Behavior Survey Results 2007, Part I: Wireline Internet Usage 1 (2007) (on file with author).

[21] *Id.* at 52.

[22] The major exceptions are Korea and Japan, where many consumers use their mobile phones to connect to the Internet and where sophisticated Web-based mobile phone services have been introduced. See Michael Fitzpatrick, *Why Mobile Japan Leads the World: A Combination of an Urban Lifestyle and Infrastructure Advantages Mean that the Fixed Internet Is Being Left Behind by the Mobile*, Guardian (London), Sept. 27, 2007, at Tech. News & Features 3.

[23] *See*, e.g., Press Release, comScore, Mobile Phone Web Users Nearly Equal PC Based Internet Users in Japan (Sept. 20, 2007), *available at* http://www.comscore.com/press/release.asp?press=1742; Press Release, Ipsos, Mobile Phones Could Soon Rival the PC as World's Dominant Internet Platform (Apr. 18, 2006), *available at* http://www.ipsos-na.com/news/act_hit_cntr.cfm?id=3049&Region=us& PDF_name=mr060418-1b.pdf; see also Darren Waters, *Why the Future Is in Your Hands*, BBC NEWS, Feb. 18, 2008, http://news.bbc.co.uk/2/hi/technology/7250465.stm.

[24] *See* Press Release, Ipsos, supra note 23. Also, cell phone unit shipments were nearly 1.2 billion units in 2007 while the number of PCs shipped globally was just over 270 million. *See* Press Release, Gartner, Gartner Says Worldwide PC Market Grew 13 Percent in 2007 (Jan. 16, 2008), *available at* http://www.gartner.com/it/page.jsp?id=584210; Ann Steffora Mutschler, *SIA: 2007 Global Chip Sales Up Modestly by 3% to $256B*, EDN, Feb. 1, 2008, http://www.edn.com/article/CA6527975.html? industryid=47037.

TABLE 1 Top 10 Properties (U.S.) January 2008.[25]

Rank	Property	Unique Visitors (in thousands)
	Total Internet Audience[26]	**184,239**
1	Yahoo! Sites	138,059
2	Google Sites	134,886
3	Microsoft Sites	119,297
4	AOL LLC	109,442
5	Fox Interactive Media	83,752
6	eBay	78,789
7	Amazon Sites	59,003
8	Wikipedia Sites	55,589
9	Time Warner – Excluding AOL	52,645
10	Ask Network	52,102

that are critical for understanding how this industry will evolve, the competitive strategies used in this sector, and their implications ultimately for antitrust policy.

Many Web businesses follow the traditional advertising-supported media model. Content is used to attract traffic. Access to that traffic is sold to advertisers. The content is usually made available for free so that advertising is the primary source of revenue and profits. Many of the leading Web properties follow this approach.

Table 1 lists the top ten Web properties in the United States based on the number of visitors to the properties' Web pages. All but three of these sites are primarily supported by advertising: Amazon and eBay are funded through transaction mechanisms while Wikipedia is funded by voluntary donations.

Google and Ask primarily sell advertising on their search results pages. AOL, Fox, Microsoft, Time Warner, and Yahoo primarily sell advertising on their various Web properties. The emergence of the advertising-supported media model for Web sites is one of the most significant changes since the dot-com bust and is the revenue source behind what is sometimes called Web 2.0.

Transaction platforms play a key role. The Web has resulted in the development of a number of transaction platforms that reduce the costs of connecting buyers with sellers and consummating

[25] Press Release, comScore, comScore Media Metrix Releases Top 50 Web Rankings for January, at tbl. 3 (Feb. 18, 2008), *available at* http://www.comscore.com/press/release.asp?press=2067.

[26] Properties include all of the sites owned by an entity, including search engines, international sites, and sites acquired. For example, the "Google Sites" include Google.com and other international homepages; search pages for specific categories such as news and images; applications such as Gmail, Google Maps and Google Earth; and sites owned by Google such as Picasa and YouTube.

trades between them. These platforms earn most of their revenues and profits from transaction fees.[27] For example, although eBay began by helping consumers sell second-hand goods to other consumers who wanted to buy those goods, it has evolved into a broad platform for connecting consumers and businesses. Other e-commerce sites, such as Amazon, have started moving from directly selling merchandise on their own behalf to providing a platform for connecting businesses and consumers.

Social networking is a critical innovation. Social networking has emerged as a new form of communication and interaction among individuals. MySpace, for example, has attracted 69 million users worldwide, who post information about themselves on the site and use it to stay in contact with friends and to make new acquaintances. A related phenomenon is that a great deal of the content on the Web is generated by users. Although sites such as YouTube, MySpace, and Facebook rely on advertising to make money as traditional media firms do, they expend few resources on actually creating or purchasing any content themselves—they focus mainly on inducing others to provide this content.

Demand- and supply-side economies of scale tend to lead to certain segments of the Web being dominated, on a national and often a global basis, by a few large firms.[28] On the demand side, some Web-based platforms create more value for each customer as they obtain more customers. A transaction platform such as eBay, for example, is more valuable to buyers because it has more sellers and more valuable to sellers because it has more buyers. Moreover, buyers and sellers appreciate being able to reach others across borders.

On the supply side, some Web-based platforms—Google, for example—also incur significant fixed costs in developing and maintaining their software platforms and in acquiring computer server and storage capacity for their activities. The average cost of providing products and services is lower for larger platforms, which can amortize these fixed costs over a larger customer base. Demand- and supply-side economies of scale result in larger firms being able to offer greater value to consumers at lower costs, as discussed in more detail below. These economies of scale do not necessarily result in a monopoly, but they do tend to limit the number of viable firms in a

---

[27] These transaction platforms usually have integrated payment platforms. eBay has PayPal, Alibaba has AliPay, and Google's commerce site has Google Checkout. In 2007, 25% of eBay's net transaction revenue came from the use of its payment platform PayPal, rather than its auction platforms and communications segment (for example, Skype).

[28] Demand-side economies of scale result when a product or service becomes more valuable as more people use it. This results from direct and indirect network effects. In the case of direct network effects, "the benefit to a network user depends directly on how many other users are hooked up to the network." In indirect network effects, "the benefit to a user arises indirectly because the number of users of the network affects the price and availability of complementary products." See Dennis W. Carlton & Jeffrey M. Perloff, Modern Industrial Organization 392–93 (4th ed. 2005). Supply-side economies of scale result when the long-run average total cost of providing a good or service falls as the quantity of output increases. Id. at 36–40.

The mere fact that a firm has a large share of a segment does not necessarily imply that it has economies of scale or network effects, and in fact, some of the shares discussed below are likely the result of the fact that these firms were just better than their rivals. However, for the reasons discussed below, it is apparent that indirect network effects, and to a lesser extent cost-based economies of scale, are important for these businesses.

segment.[29] In some segments, it is in fact unclear how many viable firms will remain and whether they will evolve towards monopoly.

Web platforms support many Web firms. These giant global Web-based firms provide platforms for other Web-based businesses. They make software services available so other businesses can provide complementary services. For example, the music service iLike.com and the online Scrabble platform Scrabulous are two of the most popular Facebook applications, raising the value of the Facebook platform as a whole. Google makes its popular mapping software available to developers who are writing applications based on Google's maps. Web platforms also provide services that these businesses rely on. Many Web-based small businesses depend on auction sites such as eBay or e-tailers such as Amazon to make sales. Most blogs rely on Google to sell ad space on their sites to advertisers who want to reach blog viewers.[30] These Web platforms are examples of the multi-sided platforms that we describe in more detail below. The Web-based businesses that rely on these platforms provide complementary products and services that make the Web platform more valuable and help drive revenue to it.

II. THE ECONOMICS AND TECHNOLOGY OF WEB-BASED BUSINESSES

A. The Economics of Multi-Sided Platforms

Many of the key businesses that have arisen on the Web are what economists call "multi-sided platforms."[31] A multi-sided platform provides goods or services to two or more distinct groups of customers who need each other in some way and who rely on the platform to intermediate transactions between them.[32] Multi-sided platforms usually lower transaction costs and thereby facilitate

[29] As firms become larger, indirect network effects and economies of scale from further expansion may diminish, and congestion and managerial economies may counter the other benefits from size. Furthermore, multiple firms may coexist in a segment despite indirect network effects and economies of scale if they can differentiate their products and services from each other.

[30] *See*, e.g., Posting of David Evans to Catalyst Code, http://www.thecatalystcode.com/theconversation/blog/2007/08/06/economics-of-the-blogosphere/ (Aug. 6, 2007).

[31] Multi-sided platforms are also called "two-sided markets" by economists, even though they are not markets—at least as markets are defined in antitrust. See, e.g., Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37 RAND J. Econ. 645 (2006). Most Web-based businesses are at least two-sided because they are transaction platforms (such as eBay, Amazon, and Alibaba), which connect buyers and sellers; social networking sites (such as MySpace and Facebook), which connect friends; or advertising-supported sites (such as CNN.com and social networking sites), which connect viewers and advertisers.

[32] *See* David S. Evans, *The Antitrust Economics of Two-Sided Markets,* 20 Yale J. on Reg. 325, 328 (2003); David S. Evans, *Some Empirical Aspects of Multi-Sided Platform Industries*, 2 Rev. Network Econ. 191 (2003); Jean-Charles Rochet & Jean Tirole, Platform Competition in Two-Sided Markets, 1 J. Eur. Econ. Ass'n 990 (2003).

value-creating exchanges. They tend to arise where there is some value from getting multiple sides together but transactions costs or other obstacles stand in the way. eBay, for example, drastically lowered the cost of exchange between buyers and sellers of second-hand goods.

Multi-sided platforms usually perform each of three interrelated core functions to some degree.[33] First, they serve as matchmakers to facilitate exchange by making it easier for members of each group to find each other. That can be for love (Matchmaker.com) or money (eBay). Second, they build communities (or audiences) because this makes it more likely that members of a group will find a suitable match. Facebook provides value in part because people are more likely to find people they want to meet and because advertisers can reach a large audience. The value of the platform grows as the audience grows. Third, they provide shared resources and reduce the cost of providing services to multiple groups of customers. This is an especially important characteristic of the software platforms discussed below.

One key feature of multi-sided platforms is the presence of indirect network effects.[34] This means that the value that a customer on one side realizes from the platform increases with the number of customers on the other side. Consumers looking to buy something value a search engine more if it provides advertisements that are more relevant to their search, while companies value advertising on a search engine higher if they are more likely to reach potential consumers.

Another key feature is that multi-sided platforms must cater to multiple, distinct customer groups simultaneously. To establish a two-sided platform, for example, the founders must solve a chicken-and-egg problem: customers on Side A will not participate without customers on Side B, but customers on Side B will not participate without customers on Side A. YouTube had to pursue people who wanted to post videos, people who wanted to watch videos, and advertisers who wanted to reach these viewers. These features make the profit-maximizing calculus for a multi-sided platform more intricate than for a traditional business. A firm operating one of these platforms must consider the demands of all sides, the interrelationships between these demands, the costs directly attributable to each side, and the costs of running the platform.

Further complicating this calculus is the fact that the profit-maximizing prices for multi-sided platforms can result in users on one side getting a price that is less than the incremental cost of serving that customer, and sometimes the price for users on one side may even be less than zero.[35] The side that is "needed more" or that is "harder to get" may receive a price break; conversely, the side that gets the most value out of access to members of the other side likely

---

[33] *See* David Evans & Richard Schmalensee, *The Industrial Organization of Markets with Two Sided Platforms,* 3 Competition Pol'y Int'l 158–59 (2007).

[34] *See*, e.g., Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects,* 8 J. Econ. Persp. 93 (1994).

[35] *See* Mark Armstrong, Competition in Two-Sided Markets, 37 Rand J. Econ. 668, 668–70 (2006); Geoffrey G. Parker & Marshall W. Van Alstyne, *Two-Sided Network Effects: A Theory of Information Product Design,* 51 Mgmt. SCI. 1494, 1498 (2005); Rochet & Tirole, supra note 31, at 659.

bears more of the cost.[36] As an empirical matter, many multi-sided platforms make their money from one side and make access to the platform available to another side for a price that does not cover the cost of provision.[37] Facebook, for example, is free to users and makes money by selling advertising.[38]

There are several major classes of industries in which most if not all of the businesses are based on multi-sided platforms. These include advertising-supported media, such as newspapers, magazines, radio, and television; payment services, such as credit and debit cards; exchanges, such as auction houses, commodity exchanges, and financial exchanges; and dating and matchmaking businesses, such as singles bars and matchmaking services. Another major class consists of industries that have software platforms as their underlying technology.[39] These include computer operating systems, mobile telephones, personal digital assistants, and video game consoles.[40] Many Web-based businesses are also included.

B. Software Platforms

A software program is a "platform" if it provides services on which other Web software can rely. Typically a software platform includes modules of code that other software programs can access through application programming interfaces (APIs). By relying on these APIs, software developers can obtain services that enable them to write software programs that are complementary with the software platform and useful to those who rely on the software platform. By relying on Facebook's APIs, Scrabulous provides a game for Facebook users and thereby makes Facebook a more valuable social networking site for those users.[41]

Historically, a major type of software platform consisted of operating systems that run on personal computers or on servers that are nodes in an organization's network of computers. Software applications such as Microsoft Word that run on operating systems are also installed on these desktop or server computers.[42] The software platforms that are central to Web-based businesses reside on servers that are attached to the Internet. Moreover, applications that work with these platforms

---

[36] *See* David S. Evans & Richard Schmalensee, Catalyst Code: The Strategies Behind The World's Most Dynamic Companies 71–92 (2007).

[37] When there are more than two sides, at least one side must make money. *Id.*

[38] *See* Facebook, About Facebook, http://www.facebook.com/about.php (last visited Apr. 23, 2008).

[39] *See generally* David S. Evans, Andrei Hagiu & Richard Schmalensee, Invisible Engines: How Software Platforms Drive Innovation and Transform Industries (2006).

[40] *See* id. at 1–2.

[41] The owners of Scrabble have objected to this take on their game. See Heather Timmons, *Scrabble Tries to Fight a Popular Imposter at Its Own Game,* N.Y. TIMES, Apr. 7, 2008, at C7.

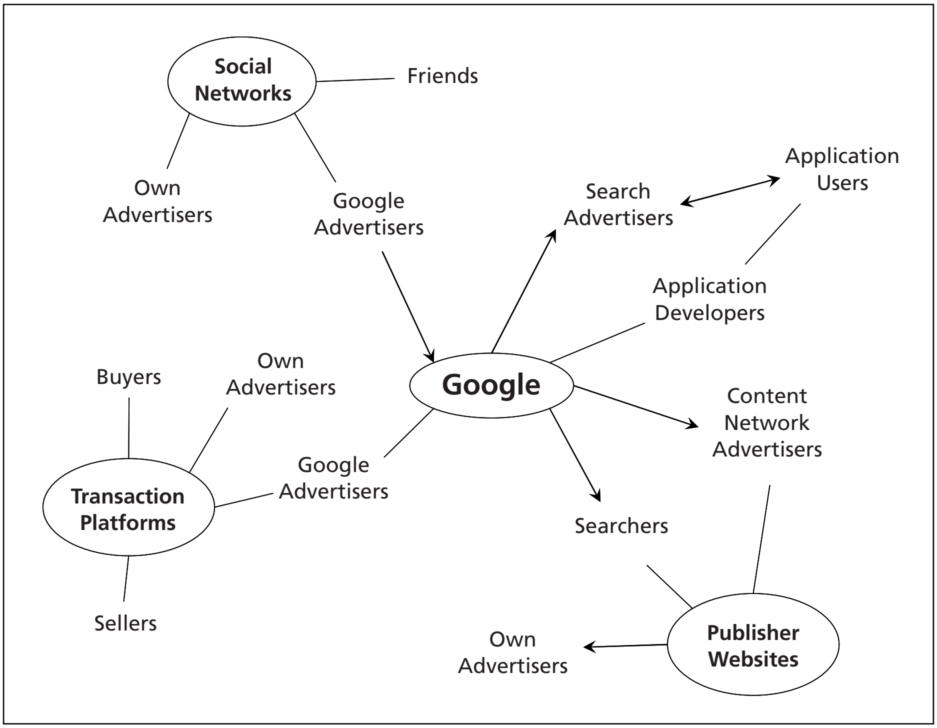
[42] An operating system is a type of software platform that manipulates the computer hardware in addition to providing code that supports other software and hardware applications.

may reside on other servers that are attached to the Internet. This has resulted in what is sometimes called "cloud computing," in which the software platform and possibly the application primarily reside on several interchangeable computers that the individual user accesses through the Internet. Google's search-based advertising platform is an example. The search engine that individuals use to conduct search queries, much of the software that advertisers rely on for advertising campaigns, and much of the software that publishers rely on for inserting advertisements into their Web pages reside on vast interconnected but indistinguishable "server farms" that Google operates around the world.

C. The Interconnected Web Ecosystems: The Example of Google

The economics and technology of Web-based businesses have resulted in an ecosystem that consists of interconnecting multi-sided platform businesses—based on software platform technology—that provide services to each other, to many other Web businesses that depend on them, and to consumers. This pattern can be seen by starting with Google's advertising platform and considering the businesses that are connected to that node. The relationships are shown in Figure 1.

FIGURE 1 Google's Advertising Platform Connections.



Google's advertising platform enables companies to insert ads based on keywords used in a search query, in which case the ad appears on the search results page, or based on the keywords found in a Web site that belongs to Google's network of Web publishers.[43] Google's search engine makes money by drawing traffic to its search results pages, where it sells and places advertising. That search engine also helps people find Web-based businesses—including publishers and e-tailers—that are not paid advertisers. Those businesses benefit from Google's search engine, but Google does not charge them for being listed in the organic search results that appear on the left-hand side of the search results page. Google also provides advertising services to Web publishers. Those publishers make space available for Google to insert ads; Google sells that space to advertisers and pays the publishers a portion of its ad revenues.

Many of the entities that Google interconnects with are also multi-sided platforms. Web publishers operate two-sided platforms in which they use content to attract viewers and sell access to those viewers to advertisers. Many small publishers, including blogs, rely entirely on Google to sell their advertising space. Many large publishers use Google to sell some portion of their advertising space; some of them also have Google search boxes and receive payments from Google for advertising revenue that results from their visitors clicking on ads on Google's search results pages. Social networking sites are similar to Web publishers in using advertising to make money. The site attracts traffic by providing social networking and makes money by selling that traffic—and data related to individual users—to advertisers through platforms such as Google. Google's advertising platform also intersects with eBay's transaction platform. eBay buys advertising on Google's search results pages to attract potential buyers to the various products and services on eBay. In addition, eBay makes advertising space available to Google and receives payments in return.

Google makes its APIs available to software developers that are writing programs to provide other services. In return, Google reserves the right to insert advertising on those services. Since January 2007, developers have written around 20,000 "gadgets" used across 100,000 Web sites. These mini-applications use the Google Gadgets API and can run on various Google platforms (e.g., Google Calendar, iGoogle, Google Desktop, Blogger, Google Maps, and Orkut). Often they can be embedded in other Web pages and run on other third-party applications (e.g., MyAOL). Developers can also create map applications on their Web sites using the Google Maps API. For example, using the Google Maps API, Orbitz added"Orbitz Updates" to its site, a map which shows real-time user-reported weather, traffic, parking, and wait-line conditions at U.S. airports.

[43] *See* David S. Evans, The Economics of the Online Advertising Industry 37 (Jan. 2008) (unpublished manuscript, on file with the Northwestern University Law Review).

D. Scale and Dominance in the Web-Based Economy

The economics and technology of Web-based businesses have resulted in the emergence of companies that have substantial shares in their categories.[44] These shares are partly the result of economies of scale in production and indirect network effects for these multi-sided platforms. Table 2 reports data for a number of countries on the two largest platforms in three representative categories: online auctions, search-based advertising, and social networking.[45] These categories are highly concentrated in every country. Moreover, the search-based advertising and online auction categories are dominated by the same firm in most countries for which data is available. eBay is the largest auction site, with over a 90% share in all of the countries for which data is available. Google is the leading search provider, with a share in excess of 80% in seven out of the sixteen countries for which data is available and a share in excess of 50% in twelve out of the sixteen countries. Social networking does not have a single leader, although the leading social networking site has more than a 50% share in most countries for which there is data.[46]

The Web economy is still young compared to other industries. Some of the leading firms are not even a decade old. It remains to be seen whether they maintain their leadership and the extent to which other platforms, through differentiation, can survive. Yahoo has long been a leading portal and advertising platform, but its market value fell 33% between January 31, 2007 and January

---

[44] I am using the term "category" to refer to commonly known types of products or services, such as social networking, portals, instant messaging, and online auctions. These categories do not necessarily correspond to markets because products and services in one category can substitute to some degree for products and services in another category. Moreover, the extent of cross-category substitution can vary over time; for example, in my experience people are increasingly using social networking sites such as Facebook as their entry point when they sign on to the Web rather than a traditional portal such as Yahoo.

[45] I have chosen these three because, as discussed below, they reflect the most important types of platforms for the foreseeable future. However, the same basic points apply to instant messaging, online payment systems, and webmail.

These categories do not necessarily correspond to relevant antitrust markets, and any analysis of market definition should properly consider the two-sided issues mentioned above. As noted below, these shares are based on categories as reported by comScore and may overstate or understate the true significance of the leading companies. First, the shares of the main search engine providers (Google, Yahoo, Microsoft, and Baidu) are understated because comScore includes searches that are done within Web sites such as eBay in its calculation of search shares, even though eBay is not generally used for Internet searches. Second, the calculation of search query shares further understates the search revenue share for Google because Google earns a higher revenue per search than Microsoft or Yahoo. See Miguel Helft, *A Long-Delayed Ad System Has Yahoo Crossing Its Fingers,* N.Y. Times, Feb. 5, 2007, at C1; Search Marketing Communications, http://cohn.wordpress.com/category/revenue-per-visit/ (Feb 7,2008).

[46] Facebook's implied market value after investments by Microsoft and Chinese billionaire Li Kashing is $15 billion; that suggests that at least some investors are betting that Facebook will become the leading social network. See Thomas R. Eisenmann & Brian Feinstein, *Facebook Platform* 1 (Harvard Bus. Sch. Case Study No. N2-808-128, 2008); Suzy Jagger, *Li Ka-shing Makes Big Impression on Facebook*, Times Online (London), Mar. 29, 2008, http://business.timesonline.co.uk/tol/business/industry_sectors/technology/article3642805.ece.

TABLE 2 Shares of Market Leaders in Major Internet Platforms.[47]

Measure	Auction Page Views		Searches[48]		Social Networking Page Views[49]	
	Leading Platform	Second Platform	Leading Platform	Second Platform	Leading Platform	Second Platform
Argentina			Google 91%	Yahoo! 4%		
Australia			Google 79%	eBay 6%	MySpace 43%	Bebo 25%
Brazil			Google 90%	Yahoo! 2%	Google 98%	Fotolog 1%
Canada			Google 78%	Microsoft 6%		
China			Baidu 54%	Google 19%		
France	eBay 99%	Delcampe 0%	Google 82%	Microsoft 3%	Skyrock 78%	Facebook 7%
Germany	eBay 99%	Yatego GmbH 0%	Google 80%	eBay 6%	StudiVZ 43%	Schueler 12%
Hong Kong					Facebook 50%	Xanga 30%
India			Google 81%	Yahoo! 11%	Google 87%	Facebook 6%
Italy	eBay 99%	Bidplaza 0%	Google 85%	Telecom Italia 3%	Netlog 27%	Badoo 23%
Japan			Yahoo 49%	Google 41%	Mixi 62%	Google 4%
Malaysia					Friendster 58%	MySpace 25%
Mexico			Google 89%	Microsoft 4%		
New Zealand					Bebo 81%	Facebook 7%
Russia			Yandex 52%	Google 32%		
Singapore					Friendster 54%	Facebook 15%
Spain	eBay 94%	Mercado-Libre 3%	Google 93%	Microsoft 2%	Fotolog 50%	Metroflog 9%
South Korea			NHN 65%	Lycos 16%	CyWorld 97%	Paran.com Club 0%
Taiwan					Microsoft 23%	Yahoo! 16%
UK	eBay 98%	Price-drop tv 0%	Google 74%	eBay 6%	Bebo 40%	Facebook 28%
USA	eBay 96%	Bidz 1%	Google 53%	Yahoo! 17%	MySpace 62%	Facebook 21%
Worldwide	eBay 57%	Taobao 14%	Google 62%	Yahoo! 13%	MySpace 17%	Facebook 15%

[47] comScore, Mymetrix Key Measures Report (Dec. 2007); comScore, Mymetrix Qsearch 2.0 Key Measures Report (Dec. 2007).

[48] The search figures are shares as reported by comScore. They include searches on Web sites where searches are primarily or exclusively used to search within the site rather than generally on the Internet. They also include searches on sites where advertising is not featured. Shares of search engines used for general searches on the Internet and shares of search engines relevant to search advertising will likely be higher than the shares reported here.

[49] Social networking shares are also as reported by comScore. They include blogging sites such as Blogger. If these sites are excluded, the shares of the market leaders would be higher.

31, 2008. Following this decline, Microsoft announced its desire to acquire Yahoo on February 1, 2008. Despite its past success and enormous user base, some analysts concluded that Yahoo could not succeed on its own.[50] Other Web giants have also encountered problems. The growth of eBay has slowed, and the company was undergoing a shakeup in management in early 2008. It faces increased competition from Amazon, Google, and other Web properties that provide transaction platforms for businesses. The capital markets have also expressed profound uncertainty over Google's growth. Its share price fell by 34% between January 2, 2008 and April 17, 2008, the day it announced its first quarter earnings.[51] Google reported a 30% increase in quarterly profits that day, and its stock increased by 23% by April 22, 2008.[52]

III. COMPETITION AND REGULATORY POLICY

Antitrust scrutiny of the leading Web-based platforms from around the world was, and is, inevitable. This scrutiny has come, so far, either through self-initiated investigations by competition authorities, through complaints by the companies' diverse stakeholders, or through complaints by rivals.[53] These Web platforms have large shares of the segments in which they operate. While one can debate whether these segments correspond to well-defined antitrust markets, the competition authorities and complainants may take these categories as a starting point. Under European Community law, a firm can be considered dominant with a share as low as forty percent.[54] Many of these platforms have segment shares that exceed eighty percent in many countries. The European Commission suggested, in its case against Microsoft, that such "super-dominant firms" should receive even greater scrutiny; some observers believe that the Court of First Instance has agreed

---

[50] *See*, e.g., Michael Liedtke, *Microsoft Woo Yahoo?,* St. Paul Pioneer Press, Feb. 12, 2008, at C1; Claudine Beaumont, *Fading Yahoo! Needs Microsoft Rescue*, Telegraph, Feb 2, 2008, http://www.telegraph.co.uk/money/main.jhtml?xml=/money/2008/02/01/bcnmicro601.xml; Grouchy Geek, http://grouchygeek.blogs.fortune.cnn.com/2008/02/01/yahoo-needs-microsofts-help/ (Feb. 1, 2008, 11:56 a.m.).

[51] *See* Google Finance, Google Inc., http://finance.google.com/finance?chdnp=1&chdd=1&chds=1&chdv=1&chvs=maximized&chdeh=0&chdet=1208462400000&chddm=28934&q=NASDAQ: GOOG&ntsp=0. In fact, Google's stock price had been stumbling since November, 2007. See Posting of Jim Kerstetter to Cnet Tech News Blog, http://www.news.com/8301-10784_3-9920376-7.html (Apr. 16, 2008, 11:06 PDT).

[52] *See* Press Release, Google, Google Announces First Quarter 2008 Results (April 17, 2008), *avail* able at http://www.google.com/intl/en/press/pressrel/revenues_q108.html; Google Finance, Google Inc., http://finance.google.com/finance?chdnp=1&chdd=1&chds=1&chdv=1&chvs=maximized&chdeh=0&chdet=1208894400000&chddm=1564&q=NASDAQ:GOOG&ntsp=0.

[53] In keeping with the global focus of this chapter, "complainants" is used to refer to parties that complain to a competition authority as well as plaintiffs in private actions, which is the dominant form of antitrust enforcement in the United States.

[54] *See* Case T-219/99, British Airways plc v. Comm'n, 2003 E.C.R. II-5917, ¶¶ 211, 223–25; Case COMP/38.233, Wanadoo Interactive, Commission Decision of July 16, 2003, ¶ 227.

with the Commission in its Microsoft judgment.[55] Under U.S. law, firms that have market shares in excess of sixty percent are often considered to have monopoly power.[56] Although it has become more difficult for plaintiffs to prevail on various monopolization theories in the United States, the caselaw on tying products together and offering discounts for bundles of products continues to provide significant opportunities for plaintiffs to pursue cases.

The existence of indirect network effects and economies of scale means these platforms are competing in "winner take all" and "a few winners take all" markets. That leads to aggressive struggles to win market share at the expense of rivals. Competition authorities worry and rivals complain, however, when tough business tactics succeed in reducing the rivals' sales, thereby foreclosing them from the market. Complicating matters, competition authorities and courts have difficulty distinguishing procompetitive from anticompetitive business practices for multi-sided platforms. For example, as mentioned previously, multi-sided platforms often charge prices that are below cost to customers on one or more sides of the platform. In some circumstances, these low prices drive out competition as a result of what business strategists refer to as "envelopment."[57] Rivals who lack the money-making side of the platform that subsidizes the money-losing product cannot survive.

Given the probable evolution of the Web-based economy, several competition and regulatory issues are likely to arise in the coming years as a result of this antitrust scrutiny.

The emergence of impregnable monopolies. There are likely to be concerns over the seeming monopolization of certain segments. It is possible that the Web economy will see a constant churning of its leading players. The fact that eBay and Yahoo have lost their once apparent impregnability is consistent with the view that dominance is fleeting. However, the evolution of the Web economy thus far is also consistent with the evolution of other industries in which it takes time for the winners to emerge.[58] If so, it is possible that a handful of firms will have near-monopoly positions in certain segments and that those positions will be protected in part by indirect network effects and the economies of scale resulting from the ability to average fixed software and hardware costs

---

[55] *See* Harry First, Strong Spine, Weak Underbelly: The CFI Microsoft Decision, Global Competition Pol'y, Sept. 28, 2007, http://www.globalcompetitionpolicy.org/index.php?id=555&action=907; Renata B. Hesse, Microsoft *and the Court of First Instance: What Does It All Mean?,* Global Competition Pol'y, Oct. 3, 2007, http://www.globalcompetitionpolicy.org/index.php?id=562&action=907.

[56] *See* Gerald F. Masoudi, Deputy Assistant Attorney Gen., Antitrust Div., U.S. Dep't of Justice, Some Comments on the Abuse-of-Dominance Provisions of China's Draft Antimonopoly Law, Address Before the University of International Business and Economics Competition Law Center Conference on Abuse of Dominance: Theory and Practice, (July 21, 2007), available at http://www.usdoj.gov/atr/public/speeches/225357.htm.

[57] *See* Thomas R. Eisenmann, Geoffrey Parker & Marshal W. Van Alstyne, *Platform Envelopment* 3 (Harvard Bus. Sch. Tech. & Operations Mgmt. Unit, Working Paper No. 07-104, 2007), *available at* http://ssrn.com/abstract=996852.

[58] *See* David S. Evans, Daniel D. Garcia Swartz & Bryan G. Martin-Keating, *The Failure of E-Commerce Business: A Surprise or Not?* 3 EUR. BUS. ORG. L. REV.1, 7 (2002); Michael Gort & Steven Klepper, *Time Paths in the Diffusion of Product Innovations,* 92 ECON. J. 630, 631, 642 tbl.3 (1982) (noting the decline in number of firms).

across larger communities. American antitrust policy recognizes that such monopoly is the reward for successful investment and innovation.[59] European Community-based competition policy views competition as the more desirable outcome, and when that is not possible, it imposes significant obligations on the dominant firm.[60]

Leveraging into adjacent markets. The structure of the Web ecosystem makes it likely that dominant firms will seek to move into related markets for complementary products or services. Because these firms are based on software platforms, it is relatively easy to add new features and services. For example, Google introduced its Google Checkout payment service in competition with PayPal by extending its software platform, integrating code into Google Product Search, and bundling Google Checkout for merchants into AdWords for advertisers. Moreover, the leading Web platforms often provide complementary services. It is a well-established economic proposition that a monopoly could make a greater profit if it also owned complementary monopolies or if it could replace these complementary monopolies with competitive markets.[61] Therefore, assuming that competition is not feasible, we would expect the dominant firms to attempt to establish monopolies across other segments. This could happen through mergers or through one dominant firm challenging another, as Google is doing with eBay.

Access to facilities. Issues involving access to other platforms and to the intellectual property that enables one platform to work together (interoperate) with another are likely to be raised when, as is the case in the European Community, the law is conducive to doing so.[62] One set of issues concerns access to a "closed platform." The Apple iTunes/iPod platform is largely closed. Apple does not encourage interoperability. Indeed, Apple seeks to prevent other music stores from making music available for iPods and other companies' devices from playing music from iTunes. Although it could obtain indirect network effects from an open strategy, it has chosen a tightly integrated

---

[59] *See* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 407 (2004).

[60] Firms that are dominant are subject to the provisos of Article 82 of the EC Treaty, which has been interpreted to prohibit various forms of tying, refusals to deal, bundled price rebates, pricing below cost, and other activities. *See* Richard Whish, Competition LAW 202–08 (5th ed. 2004) (discussing examples of abuse of dominance). In many of these cases, the practices are essentially prohibited per se if the firm is dominant—has a share of an antitrust market that is higher than forty percent or so. *See* Case T-219/99, British Airways plc v. Comm'n, 2003 E.C.R. II-5917, ¶¶ 211, 223–25; Case COMP/38.233, Wanadoo Interactive, Commission Decision of July 16, 2003, ¶ 227.

[61] *See* Michael A. Salinger, *Introduction to Chapters VII and IX of Augustin Cournot,* Researches into the Mathematical Principles of the Theory of Wealth, 4 COMPETITION POL'Y INT'L 274, 280–82 (2008) ("Today, the fundamental distinction between horizontal and vertical effects is widely accepted by antitrust practitioners.").

[62] The OpenGroup, a consortium that aims to facilitate interoperability, explains that interoperability is the ability to both exchange information and to use it. "Without a way to exchange information . . . high-tech systems literally can't communicate with each other. And, if they can't communicate, they can't work—interoperate—with each other." Open Group, Interoperability Matters, http://www.opengroup.org/bus_area/interoperability/info1/IBinfo1.htm#what (last visited Apr. 10, 2008).

software-hardware business model. This issue is central to the recent European Community cases against Apple.[63] Other platforms close themselves in particular dimensions. Facebook, for example, does not allow search engines to crawl its Web site, and as a result, the content on this social networking site is not available to searchers.

Another set of issues relates to portability of data. Web platforms derive benefits from the data they collect in a variety of ways. eBay's "Feedback Forum" provides quality information on sellers that is valuable to buyers. Users input a great deal of valuable personal information into social networking sites such as MySpace and Facebook. Google retains data on search queries that it can use to refine searches and deliver ads. In these cases one could imagine competitors seeking access to this information under an "essential facilities" theory under European Community law.[64] One could also imagine competition policy cases over restrictions that prevent users from exporting their data to competing sites. This battle has already begun, as Facebook rivals are currently lobbying publicly for the portability of social networking data.

Tying and bundling. As the current spate of cases suggests, it is probable that leading Web platforms will face complaints over tying of various forms.[65] It is also a common business strategy for software platforms to expand by adding features. They face low marginal costs in doing so, they can sometimes provide efficiencies by integrating features together or making it easier for consumers to obtain them more conveniently, and they can aggregate demand over users who may value one feature but not another.[66] The D.C. Circuit Court of Appeals found that it was appropriate

---

63 *See* Eric Bangeman, EU Commissioner Criticizes iPod-iTunes Tie-in, Ars Technica, Mar. 11, 2007, http://arstechnica.com/news.ars/post/20070311-eu-commissioner-criticizes-ipod-itunes-tie-in. html; Keith Regan, *iTunes Draws Ire of EU Commissioner,* Macnewsworld, Mar. 13, 2007, http://www.macnewsworld.com/story/56255.html.

64 "[T]he essential facilities doctrine imposes liability when one firm, which controls an essential facility, denies a second firm reasonable access to a product or service that the second firm must obtain in order to compete with the first." Alaska Airlines, Inc. v. United Airlines, Inc., 948 F.2d 536, 542 (9th Cir. 1991). For a review of the implementation of the essential facilities doctrine by the European Union, see James Turney, *Defining the Limits of the EU Essential Facilities Doctrine on Intellectual Property Rights: The Primacy of Securing Optimal Innovation,* 3 Nw. J. Tech. & Intell. Prop. 179 (2005).

65 For example, in September 2007, the European Court of First Instance upheld the European Commission's accusations against Microsoft of illegal tying of the Windows operating system and Windows Media Player. See Case T-201/04, Microsoft Corp. v. Comm'n, 2007 WL 2693858 (Sept. 17, 2007). Further, in January 2008, the Commission opened an investigation into the alleged tying of Internet Explorer and other software products to the Windows operating system. See Press Release, European Comm'n, Antitrust: Commission Initiates Formal Investigations Against Microsoft in Two Cases of Suspected Abuse of Dominant Market Position (Jan. 14, 2008) *available at* http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/08/19. The Commission has also censured Apple for its tying of the iPod music player and the iTunes music store. *See* Bangeman, *supra* note 63; Regan, *supra* note 63.

66 *See* Evans, Hagiu & Schmalensee, supra note 39, at 303–36. Demand aggregation is most easily seen with newspapers: many people only read portions of the newspapers; however, by offering various features, readers find enough content to persuade them to subscribe to the newspaper.

to apply a rule-of-reason legal standard to tying for software platforms[67] because of the possibility that there were particularly compelling efficiency explanations for the practice.[68] Whether the U.S. courts treat the Web-based companies as software platforms, however, remains to be seen; these companies have different business models and practices than Microsoft did with Windows, which was the subject of the D.C. Circuit decision. Moreover, the European Community's Court of First Instance has reaffirmed the Community's formalistic approach to tying in the Microsoft judgment.[69]

Envelopment and predation. Multi-sided platforms are a bit like clumsy giants stepping on other creatures as they move through the ecosystem. Although they may crush competitors intentionally, this may also happen as a natural byproduct of legitimate pricing and design decisions. Multi-sided platforms—and this is particularly true with the leading Web platforms—give many features and services away, often for the purpose of attracting traffic. They can readily crush companies that charge for features and services that they offer for free. But it is not only the little guy that faces concern from this strategy. Google Checkout can undercut eBay's PayPal because Google, unlike eBay, can obtain more advertising revenue from having an efficient payment method and secure data that it can use to target ads better. Likewise, Google's ability to subsidize software with advertising poses a threat to Microsoft and other software companies that charge for software; Microsoft is moving rapidly into online advertising just so that it can have a source of revenue similar to a key rival. Nevertheless, one would expect that tying, bundling, and pricing strategies that foreclose rivals will lead to competition policy investigations and prosecutions.

IV. CONCLUSION

At the inception of most new industries, hundreds of firms enter.[70] They battle it out over time. A few winners usually emerge—firms that have secured economies of scale or that have benefited from superior management or both. This pattern has been repeated numerous times over the course of the second industrial revolution that started after the Civil War. Consider the automobile industry. It began with the invention of the automobile by Karl Benz in 1885. The first commercial automobile company in the United States was Duryea, which entered the market in 1893. As of 1908, 253 automobile companies were competing in the U.S. This was whittled down to the big

[67] "A rule-of-reason analysis requires first checking whether the facts of a given case suggest that anticompetitive tying is a possibility and then weighing those anticompetitive effects with the benefits resulting from a tying policy." See David S. Evans, A. Jorge Padilla & Michele Polo, *Tying in Platform Software: Reasons for a* Rule-of-Reason *Standard in European Competition Law*, 25 World Competition 509, 514 (2002).

[68] *See* United States v. Microsoft Corp., 253 F.3d 34, 93 (D.C. Cir. 2001).

[69] *See* Christian Ahlborn & David S. Evans, *The Microsoft Judgment and Its Implications for the Competition Policy Towards Dominant Firms in Europe,* 75 Antitrust L.J. (forthcoming 2008).

[70] *See* Gort & Klepper, *supra* note 58, at 631.

four—General Motors, Ford, Chrysler, and American Motors—by 1960. Many other industries followed a similar course.

Antitrust scrutiny often follows consolidation. The leaders in the industry have large market shares, which, under longstanding antitrust practice, makes them vulnerable to claims of unlawfully maintaining or acquiring a monopoly or running afoul of other antitrust laws that have a market power screen. AT&T was hit with its first major antitrust case in 1911, twenty-six years after it opened the first telephone exchange. IBM faced its first antitrust case forty-seven years after it received patents for the punch card machine.

We can expect the Web-based industries to follow the same trajectory, and thus far they have. Massive entry has taken place. As with many new industries, we remember the YouTubes that succeeded but we forget that Google Video and hundreds of other start-ups tried and quickly failed. There are some differences, however, which suggest that antitrust controversies will arise more quickly than they did for other industries. The first is speed. Although the notion of "Internet Time" may have been exaggerated, it is true that Web-based firms can achieve leading positions in many countries around the world very quickly. The second is complexity. Almost all of the leading Web-based firms have intricate multi-sided business models. The third is interconnectedness. The Web economy is interconnected, which leads to dependencies and rivalries that can create conflict and antitrust complaints.

As a result, the competition authorities and courts will have a challenging set of issues to deal with concerning the Web-based economy in the years to come. The future will bring merger cases, as firms seek to consolidate to achieve economies of scale and indirect network effects; refusal-to-deal cases, as closed platforms deny others access to their communities; predation cases, as rivals complain about free offerings that foreclose or destroy them; tying cases, as platforms use software platform technologies to add features and functions, which in some cases will foreclose their rivals; and exclusive dealing cases, as platforms lock up traffic to achieve indirect network effects. Courts and competition authorities should exercise care in balancing the need to protect long-run social welfare against the need to stop anticompetitive strategies in this highly dynamic and complex part of the economy.

Part Four

PAYMENT CARDS

CHAPTER

11 More than Money

ABSTRACT

The simple payment card has been around since at least the beginning of the twentieth century. Hotels, oil companies, and department stores issued cards before World War I. In response to customer requests, Sears began offering lines of credit in 1910 to customers of "unquestionable responsibility," although the Sears card came more than a decade later. Some large retailers gave cards to their wealthier customers that identified them as having a charge account with the store. By the 1920s, several department stores allowed cardholders to pay off their bills in monthly installments. Metal "charge-plates" with embossed consumer information were introduced by department stores in 1928. During the 1920s as well, oil companies issued "courtesy cards" for charging gas. By the end of World War II, charge cards were no longer a novelty, but they were about as far from the cards of today as barter was from coin.

"Perhaps you would like to see what our credit cards are like," [Doctor Leete asks of his guest, Julian West]. "You observe...that this card is used for a certain number of dollars....The value of what I procure on this card is checked off by the clerk, who pricks out of these tiers of squares the price of what I order." [Mr. West, from Boston, has awakened 113 years in the future, in the year 2000.]

—Edward Bellamy, Looking Backward, 2000–1887 (1888)

I. DINING ON THE CUFF

Restaurants did not issue cards. In 1949, Frank McNamara, the president of a New York credit company, was having lunch in Manhattan. A year later, as we mentioned earlier, he had a thriving business based on this experience. He was written up in *Newsweek* two years later: "Halfway through his coffee, McNamara made a familiar, embarrassing discovery; he had left his wallet at home. By the time his wife arrived and the tab had been settled, McNamara was deep in thought. Result: the 'Diner's Club,' one of the fastest-growing service organizations." (This is the earliest rendition of the story we've found. One journalist several years later gave the credit to Alfred Bloomingdale, then the president of Diners Club, and changed the meal to dinner.)

Following McNamara's epiphany, people began carrying charge cards in their wallets. McNamara and an associate, Ralph Schneider, started small. Beginning with $1.8 million of start-up capital, they signed up fourteen New York City restaurants and gave cards away to selected people. By the card's first anniversary there were 42,000 cardholders, each paying $21 a year for membership in the "club." And 330 U.S. restaurants, hotels, and nightclubs accepted these cards; they paid an average of 7 percent of the cardholder's bill to Diners Club. In March 1951, Diners Club handled $3.5 million of exchanges between cardholders and merchants, and reportedly made about $70,000 in pretax profit. At that pace, it was handling $41.5 million in transactions annually. Unlike store cards, Diners Club cards provided a broader medium of exchange—one that extended to at least all the merchants in the club. And it was more than money because consumers didn't have to pay right away. (As always, we have adjusted figures to 2008 dollars. In this chapter we sometimes round figures and use "about" or "around" to signify that we've done so. We report a few figures from 2005 and later and these we do not adjust.)

Figure 1 Frank McNamara, the man who invented the modern card industry

Figure 2 One of the early cardboard Diners' Club cards

That club expanded rapidly. In 1956, it had an annual transaction volume of more than $290 million. The card was accepted at nine thousand establishments, according to the *New York Times*, from Anchorage to Tahiti. By then its merchant coverage had expanded beyond restaurants to auto rental agencies and gift shops—almost the gamut of travel and entertainment locations. Two years later, Diners Club had an annual charge volume of more than $465 million, and earned gross profits of $40 million from merchant discounts and cardholder fees.

McNamara and Schneider had not only discovered the idea of a general-purpose payment card; they had also discovered a pricing strategy that got both merchants and cardholders on board, one which has been followed by payment card systems since. By 1957, Diners Club had raised the cardholder fee to $30, but had left the merchant fee at 7 percent. It nevertheless continued to earn most of its revenues, about 70 percent, from merchants.

Diners Club faced competition soon after its entry. Information is spotty on some—National Credit Card, Inc., for instance, started its card program in 1951, operated in forty-two states, but had filed for bankruptcy by 1954. A 1955 *Newsweek* article referred to Trip-Charge with eighty-five thousand cardholders and nine thousand merchants after a year in business; it asserted that the founder, Sidney J. Rudolph, "is now a hairbreadth from realizing his hopeful company motto: 'Charge Everything Everywhere.'" Esquire and Duncan Hines both had travel and entertainment cards, which merged in 1957. *Gourmet* magazine had a club for diners, too.

Merchant coalitions were another source of competition. Hotel owners balked at the fee they had to pay on charge cards. In 1956, the American Hotel Association established the Universal Travelcard. It didn't charge participating hotels and rental car agencies any fee but billed cardholders the same $30 fee as Diners Club. The National Restaurant Association, with sixty thousand members, signed on. One might wonder how they could get by without the 70 percent of the pie from merchant fees that Diners Club received. Part of the answer is that they didn't do central billing. Each hotel billed its customers directly, although the card association did ensure payment. There was a railroad card and an airline card as well.

Some banks had also entered the payment card business by the 1950s. But their cards—sometimes called "shopper" cards—targeted a different cardholder and merchant base. These bankcards were typically held by "housewives," as the newspapers of the day put it, and could be used only at retail stores in the locality the bank did business. Franklin National Bank started one of the first shopper cards on Long Island in 1951, and one hundred or so banks—mainly small, suburban banks in the Northeast—followed. While one hundred may seem like a large number, one should keep in mind that in 1951, there were 13,455 commercial banks as well as many more credit unions and savings and loans. Banks generally charged retailers the then-standard 5 to 7 percent merchant fee. Cardholders reportedly didn't pay any direct fees and, as with the other card programs, were supposed to pay their monthly bill in full. These cards were thus charge cards, though the press of the day called them credit cards.

Of the early major entrants, only Diners Club survived the decade as a stand-alone company. It bought Trip-Charge in 1956 and Esquire's card program in 1958. The bankcards failed mainly because they had trouble signing up merchants, and only twenty-seven of the shopper cards were still in existence by 1957. The Universal Travelcard and the Gourmet Magazine Club card were swallowed in 1958 by a new competitor that appeared during what became a critical year for the future history of the card industry.

Diners Club also expanded overseas in the mid-1950s, using franchise agreements to extend its reach to Europe. As in the United States, European hotel trade associations posed strong resistance to the travel and entertainment (T&E) cards, going so far as to expel members who accepted payment cards that required a merchant discount. Some hotels in England and Switzerland chose to flout the prohibition; they accepted the T&E cards and formed their own association. They also went a step further: the newly formed hotel association created the BHR credit card in the 1950s, which evolved into the EuroCard in the mid-1960s.

II. THE CLASS OF 1958

Though planning had started years earlier, several competitors rolled out new cards in 1958. In September, Bank of America started a credit card in California. In October, American Express launched its national charge card, and Hilton Hotels spun off its hotel card into Carte Blanche.

A. Bank of America

California did not have restrictions on branch banking in 1958. With an economy larger than Japan's, California was able to support several large banks. Bank of America was the largest, with over six hundred branches throughout the state. It had started as Bank of Italy in 1904, founded by one of the greats of U.S. banking, A. P. Giannini. By 1958, Bank of America was the largest bank in the United States.

Despite its size, though, Bank of America was cautious about offering a payment card, even though one of its small competitors, First National Bank of San Jose, had started a credit card in 1953. Bank of America considered introducing its own card in 1954, but initially decided that there wasn't a good enough business case. After studying the emerging industry over the next few years, it decided to introduce a credit card in 1958. Creditworthy customers would receive cards with limits of either about $1,700, or $3,000; prior authorization would be required for purchases over about $300; and a revolving credit option was available for some cardholders. Revolving credit was the innovation that distinguished this card from existing charge cards.

The bank conducted a market test in Fresno, California, in fall 1958. Three hundred retailers signed up initially, and every Bank of America customer in the Fresno area received a card. According to one study, "This mass mailing of 60,000 cards had been William's [the executive in charge of the effort] solution to the problem of how to convince retailers that enough individuals would possess a card to make their participation in the program worthwhile. His solution worked, for during the next five months another eight hundred Fresno-area retailers joined the newly named 'BankAmericard' program." Bank of America had planned to track the financial results of the card in Fresno before going statewide, but fear of competition persuaded it to accelerate the launch. It expanded throughout the state during the following year. By the end of 1959, twenty-five thousand merchants accepted the card and almost two million California households had one.

Things did not go well at first: fraud was rampant, the number of delinquent accounts was five times higher than expected, large retailers resisted joining, and echoing an old theme, "Public criticism came from those who viewed credit as a societal evil." The program lost $52.6 million in 1960. The bank worked on collection problems and reduced the merchant fee to as low as 3 percent to entice retailers. Delinquencies declined and the merchant base increased to thirty-five thousand in 1962. The card turned its first operating profit in 1961.

B. American Express

American Express started as an express mail company in 1850. Money was one of the things people wanted to move around the country, especially after the post–Civil War expansion of the rail network created national markets. Of course, people also wanted their money delivered safely. The U.S. Post Office developed the money order; American Express introduced a competing product. Both products were subject to theft, and neither was a good substitute for cash.

The travelers cheque, invented by an American Express employee and offered beginning in 1891, was a significant advance over the money order. The cheques came in multiple denominations just like cash and had the dual-signature system (sign when you obtain, and sign when you cash) that remains the major security device to this day. Initially, people could cash travelers cheques only at American Express offices; later, they could cash them directly at merchants. The major selling point for consumers was security: American Express guaranteed payment, but it also assumed responsibility for lost or forged checks. The product was a highly profitable one for American Express for over a century, although its popularity has declined steadily over the last three decades. The profits all came from individuals. A consumer purchasing $500 in travelers cheques, say, would pay American Express a fee in addition to the $500, and American Express would continue to earn interest on the amount invested in the cheques until they were cashed. Stolen or misplaced cheques that were not cashed or replaced would be pure profit.

A hundred years after its formation in Buffalo, New York, American Express was the world's largest travel agency and operated the world's largest private mail service. Between its cheques and travel offices, it was profiting enormously from the boom in international travel following the end of the Second World War. The number of American Express travel agencies grew from fifty at the end of the war to nearly four hundred ten years later. The company sold approximately $6.5 billion worth of travelers cheques in 1951, and by the end of the decade claimed to control 70 percent of the U.S. travelers cheque business.

The Diners Club charge card was a new competitor for American Express. By 1953, American Express had begun planning its response. It considered buying Diners Club in 1956, but rejected the idea. The company finally entered the charge card industry on October 1, 1958, with 17,500 merchant locations and 250,000 cardholders. It achieved this scale quickly by buying the Gourmet Magazine Club card and the Universal Travelcard. Within seven months of launching its card operations, American Express had over 600,000 cardholders. By late 1960, it had a charge volume of over $584 million and 750,000 cardholders.

American Express adopted a slightly different pricing policy than Diners Club. It initially set its annual fee $1 higher (in 1958 dollars) than Diners Club's $5, thereby suggesting that it was the more "exclusive" card (in 2008 dollars, the American Express annual fee was $42 while the Diners Club fee was $35). But it set the initial merchant discount slightly lower than Diners Club's 7 percent: 5 to 7 percent for restaurants, according to their sales volume; and 3 to 5 percent for the recalcitrant hotel industry, depending on the hotel guest's charge level.

American Express struggled at first. Even a charge card involves extending credit for a time, and American Express, unlike the founders of Diners Club who have been in the consumer credit business, had no experience doing this. By 1961, with losses mounting, it considered selling the business to Diners Club, but decided that such a sale might not pass muster with the Justice Department. Instead, American Express hired George Waters, later known as the "Father of the Card," to run its card operations. Waters started putting pressure on customers who had not sent their payments in on time. And he raised the annual fee to $46, and later to $56. Despite the increased fees, the American Express payment card system continued to grow. By the end of 1962, there were 900,000 American Express cardholders who could use their cards at 82,000 merchant locations. In 1962, almost four years after its launch, the card operation posted its first (small) profit.

Today, American Express accounts for 13 percent of the dollars transacted on payment cards. By that measure, it is almost a quarter of the size of the Visa system (see Table 1.1 in chapter 1) and almost one-half the size of MasterCard. Diners Club, on the other hand, shrunk to almost nothing within the United States. The charge card pioneer spread itself too thin in the 1960s and early 1970s in an attempt to counter the inroads made by American Express. In particular, Diners Club tried unsuccessfully to follow the American Express model by expanding into travel clubs and travel agencies, but it lost money on both endeavors and lost sight of the newly emerging competition from the bankcards.

C. Carte Blanche

Hilton Hotels had issued a million charge cards for use at its worldwide hotel chain. After a failed attempt to buy Diners Club, it rolled its hotel card into a new general-purpose card company, the Hilton Credit Corporation, which distributed the Carte Blanche card. In 1958, it entered with a low merchant fee, 4.5 percent, which soon dropped to 4 percent. The 600,000-member National Restaurant Association, which had been complaining about the 7 percent fees charged by Diners Club and American Express, threw its official support behind the card. Nonetheless, as a result of issuing cards to the wrong people and an inefficient billing system, Carte Blanche became known in the trade as “Carte Rouge” for its steady losses.

D. What Happened to the Class of 1958

In 1960, a decade after the birth of the general-purpose payment card, there were three major national card systems. Diners Club, with 1.1 million cardholders, was still the biggest, but it faced competition from recent entrants American Express and Carte Blanche. These three, in turn, faced regional competition from BankAmericard in California and Chase Manhattan in New York City, though nothing else of significance. Chase Manhattan sold its card program to a subsidiary of American Express in 1962. This became the Uni-Card—a credit card that was available in the Northeast United States. It was sold back to Chase in 1969. Chase joined the BankAmericard association in 1972 and converted its cards to the BankAmericard brand.

Carte Blanche was sold to Citigroup in 1965. (Citigroup started as First National City Bank.) Pressured by an antitrust suit brought by the Justice Department, Citigroup sold Carte Blanche in 1968. The Justice Department was concerned that with Carte Blanche, Citigroup might limit development of the "Everything" credit card program it had started in 1967. When Citigroup quickly dropped the Everything card and had not introduced a replacement by 1978, however, the Justice Department relented, and Citigroup bought Carte Blanche back again. By that time, Carte Blanche's share of credit card volume had declined to less than 1 percent.

American Express moved past Diners Club to become the industry's volume leader in 1966. Diners Club continued to decline throughout the 1960s, in part because it lacked the travel offices that American Express used to distribute its card during the industry's early days. American Express also had a better T&E brand as a result of its travel offices and travelers cheques. Diners Club attempted to meet American Express head-on through travel and reservation system acquisitions, but it failed to make those profitable. Diners Club was sold to Citigroup in 1981. After the sale, Diners Club shifted its focus to affluent business travelers, trying to follow American Express's successful up-market strategy. The final indignity for this pioneer came in 2008 when Discover, the newest and smallest of the payment networks, bought Diners Club International for $165 million (Citigroup retained the franchise to issue Diners Club cards in the US but then sold even that to Bank of Montreal in 2009.)

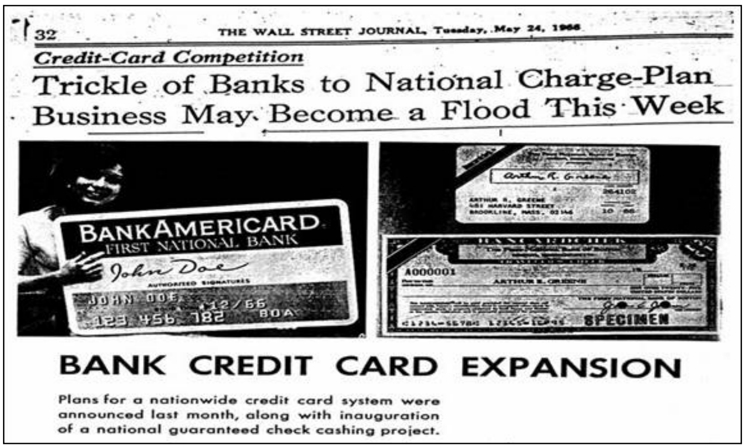
And the statewide BankAmericard became the worldwide Visa card.

III. 1966 AND THE BIRTH OF CO-OPETITION

Another watershed year for the emerging payment card industry was 1966, which marked the start of a battle between three competing business models for operating payment cards.

American Express, Carte Blanche, and Diners Club were mainly used for travel and entertainment, and thus became known as T&E cards. They did not offer credit beyond the time it took to get cardholders their monthly bills, which had to be paid in full. Nor was there a link to cardhold-

Figure 3 Bank Associations Entry in 1966


32 THE WALL STREET JOURNAL, Tuesday, May 24, 1966

Credit-Card Competition

Trickle of Banks to National Charge-Plan Business May Become a Flood This Week

BANK CREDIT CARD EXPANSION

Plans for a nationwide credit card system were announced last month, along with inauguration of a national guaranteed check cashing project.

ers' checking accounts. Many business travelers and wealthy households had one of these cards, but most Americans didn't. Data for 1966 are not available, but even by 1970 only 9.2 percent of households had one of the T&E cards.

Interstate banking regulations and other hurdles made it difficult for Bank of America to compete head-to-head with the three T&E cards. To take its card national, the California bank decided to franchise. In 1966, it announced that it would license its BankAmericard program to selected banks across the country. Each bank would operate the program independently using the BankAmericard name; merchants signed up by the franchisees would have to accept all BankAmericards, allowing consumers to use their BankAmericards at any participating merchant. Bank of America charged the franchisees a royalty of up to 0.5 percent of cardholder volume and an entry fee of about $132,000.

Unlike T&E cards, bankcards did not charge cardholders membership fees, earning revenue from finance charges and merchant discounts instead. For example, the Chase Manhattan Charge Plan, introduced in 1958, charged cardholders 1 percent of the revolved (that is, unpaid) balance every month, while charging merchants 2 to 6 percent of sales depending on volume.

The BankAmericard franchise was not limited to the United States. Major banks in countries such as Canada, Columbia, Italy, Japan, Mexico, Portugal, Spain, the United Kingdom, and Venezuela signed up as international BankAmericard franchisees around the same time as the domestic franchise system launch in 1966. In 1968, MasterCard also expanded internationally by forming alliances with EuroCard, the European card association mentioned earlier, and Banco Nacional in Mexico. The alliances allowed the MasterCard network and foreign networks to interoperate, but preserved each card as a distinct brand. In addition, MasterCard expanded in Asia by gaining member banks in Japan.

In the United States, within two months of the Bank of America franchising announcement, American Express, Carte Blanche, and Diners Club responded by offering their own franchise opportunities to banks. The American Express bankcard differed from the standard card: it offered a minimum $9,000 line of credit. American Express would split revenues with the banks: banks got a commission for signing up cardholders and revenues from credit provided by the card; and existing American Express cardholders could be converted to the bank program. American Express didn't charge additional franchise or licensing fees. Carte Blanche priced its franchise at $45 for every $5.3 million in bank assets, with a $26,429 minimum fee. Diners Club offered to franchise its card for that same fee to banks with less than $5.3 billion of assets and for $52,858 for banks with $5.3 billion or more. While the historical evidence is sketchy, it does not appear that the various efforts at franchising these cards attracted any takers within the United States. (American Express also planned to franchise its Uni-Card credit card across the country in 1968. It had a million cardholders and eighteen thousand merchants in New England, New York, New Jersey, and Pennsylvania. Instead of following through on those plans, though, it sold the Uni-Card program back to its original owner, Chase Manhattan, in 1969, and Chase converted these cards to BankAmericards in 1972.)

For many banks, there were significant negatives to the franchise system. Major banks, including Wells Fargo in California and Chase Manhattan in New York, were not eager to sign up to issue someone else's card. The successful franchise systems we're familiar with—McDonald's, the Athlete's Foot, or Mail Boxes Etc.—typically involve a prominent brand name with outlets operated by unknown local entrepreneurs. Although some franchisees can become quite successful with multiple locations, they generally have little ability or desire to promote their brand name over the franchisor's. This was not the case with the major banks.

Developing a proprietary card system was another option for banks. As we mentioned, Citigroup, in addition to owning Carte Blanche, had started its proprietary Everything card in 1967. Because Citigroup held a national banking charter and had customers across the country who were potential payment cardholders, it initially hoped to develop the Everything card into a national brand. Other banks found this option unattractive. While a national charter was, legally speaking, not necessary to issue credit cards around the country, some banks were reluctant to expand out-of-state.

Many banks found the answer in co-opetition. Banks ***competed*** for merchants and cardholders; banks ***cooperated*** at the card system level by setting operational standards. Despite the dismal experience of the 1950s, many banks decided to start cards in the 1960s. They compared the problems of going it alone to the benefits of cooperation. The calculus led them to form associations. Five banks in Illinois founded the Midwest Bank Card. By January 1967, nearly six hundred banks in Illinois, Indiana, and Michigan had joined; some of these issued one of the five original members' cards. There were also two Michigan associations. Other banks across the country followed. Three New York City banks started the Eastern States Bankcard Association in June 1967. The state and local banking groups began to develop ties with other groups. The Interbank Card Association started in 1966. Early on it included banks in Buffalo, Pittsburgh, Milwaukee, Seattle, and Phoenix. At the same time, several banks in California started the Western States Bankcard Association, issuing cards under the Master Charge service mark. By 1967, the California banks issuing Master Charge had joined the Interbank Association. By February 1968, Interbank had 286 banks in at least seven states.

It became apparent during 1968 that two competing national networks of banks had emerged: the BankAmericard franchise system, and the Interbank cooperative system. Banks—and groups of banks—started aligning with one or the other. "Just about every bank in the card field," said *Business Week*, "is convinced that it must join one or the other network." Bankers Trust in New York City went with BankAmericard and franchised the card to other banks in the New York area. Meanwhile, Citigroup converted its Everything card to Master Charge and joined Interbank. Chemical and Manufacturers Hanover joined that association as well. For the most part, the larger banks had chosen Interbank over BankAmericard. In contrast to the BankAmericard franchise model, Interbank charged only a "modest" entrance fee and a small annual fee to cover the operating costs of the joint enterprise. And as noted, banks would be selling a brand they jointly owned, rather than that of another bank. This was an important point for banks that harbored hopes of future national expansion when interstate banking restrictions were lifted—though in hindsight, that was still more than three decades away.

BankAmericard was not doing too badly by many measures. Under its franchise model, it had about 27 million cardholders and about 565,000 merchants by 1970, a sizable jump from 1.8 million cardholders and 61,000 merchants in 1966. But it was in the process of being overtaken by Interbank, which had attracted most major banks. And the franchisees were restless.

The franchisees quickly went from restless to rebellious. They had grown in importance to the system and wanted a voice in its future. In 1970, faced with this revolt as well as with operational problems, Bank of America agreed to convert the system into a membership-owned corporation: National BankAmericard, Inc. (NBI). NBI wasn't an ordinary stock corporation; instead, its members had voting rights that couldn't be bought or sold. Initially, NBI had 243 charter members, including Bank of America, Bankers Trust, and First Chicago as well as numerous smaller community and regional banks.

The brief period 1966–1970 turned out to be critical for the future development of the payment card industry. Three alternative business models battled against each other. The go-it-alone model had been the one used by American Express, Diners Club, and Carte Blanche for charge cards. Bank of America adopted this model in California, Citigroup tried it with its Everything credit card, and American Express tried it with its Uni-Card credit card. Bank of America and the three T&E card companies tried the franchise model. Finally, the co-opetitive model was used by Interbank and many other associations of banks across the country. The web of interstate and branch banking restrictions played a crucial role in the battle among these models.

By the end of the 1960s, the franchise model was dead in the United States. And the go-it-alone companies and the co-opetitives had gone in different directions. Sticking to what they knew best, the go-it-alones decided not to issue credit cards. American Express had ventured into credit with its Uni-Card, but decided to get out and did not try again for twenty years. The co-opetitives, on the other hand, focused on issuing cards with a revolving line of credit. (Debit cards were soon added, but took until the mid 1990s to have an impact.)

The basic idea behind co-opetition was clear as early as the Midwest Bank Card. The five founding members competed with each other for cardholders and merchants in the Chicago area. They cooperated in two related respects. They agreed to make their systems "interoperable." A First National Bank of Chicago cardholder could use her card at every merchant who had signed up with any of the five banks. A Harris Trust Company merchant could accept as payment any card from these banks. What the banks lost in helping their competitors, they more than gained in making their own card more appealing to cardholders and merchants. Interoperability forced cooperation in another way. When a First National Bank of Chicago cardholder bought something at a merchant who was affiliated with Harris Trust Company, Harris Trust had to get reimbursed by First National. These banks faced the same problem as in the BankAmericard franchise: system cooperation was essential to process the slips of merchant receipts that were growing exponentially with the number of participants in the system.

The two national associations encountered similar issues to Midwest Bank Card, only on a grander scale. MasterCard—which started out as the Interbank Card Association in 1966, changed its acceptance brand to Master Charge in 1969, and finally became known as MasterCard in 1979—and Visa—which began as BankAmericard in 1958, switched to National BankAmericard, Inc. in 1970, and settled on Visa in 1976—took cooperation further than any of the regional associations. First, they established rules and processes for settling transactions. Part of this involved how the merchant and the cardholder banks divvied up the transaction proceeds. Some of the regional cooperatives had initially exchanged at par so that the cardholder's bank reimbursed the merchant's bank for the entire transaction and the cardholder's bank didn't get any of the merchant fees. (The banks may have simply been applying the check model of par exchange, which they soon found unsatisfactory for cards.) MasterCard and Visa both settled on an interchange fee—a percentage of each transaction that the merchant's bank gave to the cardholder's bank.

Another area of cooperation was on the card brand. The banks that belonged to Interbank decided early on to use the Master Charge brand. That—and not the individual bank's name—was most prominent on the cards from the late 1960s and early 1970s. Visa replaced BankAmericard as the brand name for that system in 1976. The co-opetitive felt that "standardizing the somewhat confusing array of blue-white-and-gold cards issued under different names in twenty-two countries around the world" would lead to greater acceptance of the card. Focusing on the system's brand involved a trade-off basic to the co-opetitive model: choosing between doing things at the system versus the member level. Overall, the co-opetitives chose to do most things at the member level.

In a few pages we will tell the story of how the bank associations died. Yet they endured for about four decades and were responsible for the remarkable growth of the payment card industry and for shaping the industry we know today.

IV. REGULATION AND STAGFLATION

American Express prospered during the 1970s. Between 1960 and 1977, real net income grew at an average annual rate of 16.6 percent. By 1977, American Express had eight million cardholders, bringing in $459 million in annual card fees. Its lead over Diners Club and Carte Blanche had widened dramatically. American Express had decided against offering a credit card and had unloaded its Uni-Card credit card. About 200,000 of its cardholders had "corporate cards"—cards that employers ask employees to use for expenses and that provide employers with detailed spending data. Sticking with charge cards, shunning credit cards, and focusing on corporate users was a profitable strategy for some time.

Meanwhile, American Express's credit card competitors—the banks that issued credit cards and the two associations to which they belonged—struggled during this decade of government regula-

Table 1 U.S. recessions, 1969–1982

Start Date	End Date	Duration in months	Peak unemployment rate percent
December 1969	November 1970	11	5.9
November 1973	March 1975	16	8.6
January 1980	July 1980	6	7.8
July 1981	November 1982	16	10.8

Note: The National Bureau of Economic Research defines a recession as "a period of significant decline in total output, income, employment, and trade."

Sources: National Bureau of Economic Research, *U.S. Business Cycle Expansions and Contractions,* <http://www.nber.org/cycles/> (accessed April 21, 2003); and U.S. Department of Labor, Bureau of Labor Statistics.

tion, volatile interest rates, and economic stagnation. The economy went through several severe recessions in the 1970s and early 1980s (see Table 1). The economy also experienced accelerating inflation, jumping from 4.2 percent in 1972 to a peak of 9.4 percent in 1981. This led to an increase in (nominal) interest rates. (Economists distinguish between nominal interest rates, which are quoted by lenders, and real interest rates, which are lower than nominal rates by the rate of inflation. Real rates adjust for the impact of inflation on purchasing power over time; nominal rates tend to vary with the rate of inflation, all else being equal.) The one-year Treasury bill rate climbed from 4.9 percent in 1971 to 14.8 percent in 1981. As interest rates climbed, state usury laws made credit card lending a bad business proposition. Banks need a spread between the finance rate they charge consumers and their own cost of funds to cover their operational costs, including fraud and defaults, and make a profit. As inflation rose, that spread generally narrowed and even became negative in states with low caps on loan rates. Banks lost money on the credit they had already made available on cards and became unwilling to extend further credit in the face of these losses. Usury laws, and a Supreme Court decision about them, helped shape the card industry in the 1970s and beyond.

A. Usury Laws

Thirty-six states had usury laws in 1982. Some had caps that were so high they didn't matter during times of normal inflation and interest rates—for instance, Georgia had a maximum of 60 percent. But others topped out at rates that made it difficult for banks, even in not too far from normal times, to make unsecured loans profitably. Arkansas, for example, had a maximum rate of 5 percent above the Federal Reserve discount rate, giving a maximum rate of about 12 percent at the time. Banks lived with their state caps by tailoring their credit card lending. Banks in states with high limits could extend credit to a wide range of people because they were able to charge finance rates that covered the inevitable defaults, late payments, and fraud. Banks in states with low limits raised their credit criteria for issuing cards and imposed higher membership fees. For example, because Arkansas imposed a low cap on consumer loan rates, banks in that state had to set their standards relatively high and offer credit to very few people in order to hold down costs. As a result, charge-off rates in Arkansas were low, but so was the use of credit cards. (Charge-offs are credit card balances that have been written off as losses for tax purposes.)

Usury laws had a significant effect on the development of a national card industry because they limited the ability of banks to market their cards on a national or even regional basis. Interest rates that were lawful in one state were unlawful in another. A bank therefore couldn't market a card nationally or regionally and capture scale economies from wide distribution. The few banks that issued in multiple states had to incur the expense of administering multiple card programs with different terms in each state. This wasn't just an issue of processing extra paper; the banks had to adjust credit standards and collection criteria according to the finance charges they could assess in each state.

A Supreme Court decision changed the rules of the game in 1978 and helped create national competition for payment cards. First of Omaha Service Corporation, a subsidiary of First National Bank of Omaha, began to apply interest rates that were legal in Nebraska, but higher than the Minnesota rate ceiling, to its Minnesota credit card customers. Marquette National Bank of Minnesota challenged this practice. The Supreme Court sided with First of Omaha. In *Marquette National Bank v. First of Omaha Service Corp.,* the Court ruled that as a national bank, First National Bank of Omaha "may charge interest on any loan at the rate allowed by the laws of the State where the bank is located." The Court also said that a bank's "location" refers to the state in which the bank is chartered, regardless of the states in which it solicits customers.

The *Marquette* decision led to three major developments. First, nationally chartered banks started issuing credit cards from states with less-restrictive usury laws. Citigroup, for example, moved its credit card operations from New York, which at the time had an interest rate cap of 12 percent on balances greater than about $1000, to South Dakota, which had raised its interest rate ceiling to 19.8 percent. Many other banks moved their operations to Delaware, which eliminated interest rate caps in early 1981. By 1987, many large banks legally resided in Delaware, including Bankers Trust, Chase Manhattan, Chemical, Manufacturers Hanover, Morgan, and Marine Midland.

Second, in an attempt to attract or retain such movable card operations, some states began to modify their usury laws. In 1980, the same year that Citigroup announced it would be moving its operations to South Dakota, the New York State Senate passed a bill that eliminated interest rate ceilings on most types of loans, except annual rates on credit cards, which remained at 25 percent. By 1988, the majority of states still had some form of an interest rate cap on credit card loans, but many had raised their ceilings.

Third, less balkanization from state credit restraints set the stage for marketing payment cards on a nationwide basis. Citigroup led the way in the late 1970s and early 1980s: it expanded nationwide through acquisitions and mass-mail credit card solicitations. Throughout the 1980s, many other banks launched national credit card campaigns as well, including Bank of America, Chase, Continental Illinois, First Chicago, and Manufacturers Hanover. By permitting nationwide competition, *Marquette* enabled issuers to realize scale economies in marketing and processing costs, thereby making payment cards more readily available to consumers across the country. Thus, even though Arkansas still caps interest rates at 5 percent above the Federal Reserve discount rate today, consumers in Arkansas have a wide range of card choices from national issuers.

B. Duality

Federal antitrust laws also had a major effect on the evolution of the payment card industry in the 1970s. Visa's rules initially prohibited MasterCard-issuing banks from issuing its cards or handling its merchant paper. In July 1971, one of Visa's charter members, Worthen Bank and Trust Company of Little Rock, Arkansas, filed an antitrust suit, claiming that Visa's prohibition amounted to an illegal group boycott. While this case was eventually settled out of court, Visa remained exposed to similar lawsuits. In 1974, it asked the Antitrust Division of the U.S. Department of Justice for a business clearance review—in effect, a letter of approval for a rule that would prohibit dual membership by card-issuing and merchant-acquiring banks. After a year of consideration, the division declined to grant clearance, citing insufficient information. Without the division's support, Visa removed all restrictions on dual membership in mid-1976. The age of what has become known as "duality" began, and members of each system rushed to join the other.

Competition among issuers increased sharply as new members scrambled to sign up consumers for their second card. From mid-1976 to mid-1977, the number of cardholders increased by 11.7 percent for MasterCard and 13.1 percent for Visa—significantly higher growth than in preceding years. Banks that were previously exclusive to MasterCard could now sign up their merchants for Visa (and vice versa) as well as compete generally for merchants. Merchant acceptance grew sharply from mid-1976 to mid-1977, by 24.3 percent for MasterCard and 17.3 percent for Visa, again much higher than in preceding years.

One year after the restriction was removed, twenty of the nation's twenty-two-largest banks that issued cards had become dual. With increased system overlap, Visa and MasterCard twice discussed the possibility of merging system infrastructures in the 1980s, but decided against the move both times. With duality, the difference between Visa and MasterCard began to blur, at least as far as consumers were concerned. Some dual members offered common agreements, billing statements, and credit lines. Some used common advertisements for their dual cards.

Over time, almost all member institutions joined both associations, and the increasing overlap in membership led to some decline in competition between the two systems. Dual membership lost some of its appeal in the late 1990s. The Justice Department, ironically, claimed in a case filed in 1998 that MasterCard and Visa were violating the antitrust laws by having dual membership. A federal district court judge disagreed. By then, though, meaningful duality had withered, as the associations sough allegiance, and in the next decade the associations were disbanded.

C. A Painful Decade for Card Issuers

The Visa and MasterCard associations had little trouble finding banks to join their still-fledgling associations in the 1970s. First, banks were afraid of being left behind. Many banks got into the card business to preempt or counter their competitors' plans. Second, there was a growing belief that credit cards were a stepping-stone to the "cashless society." Third, cards created an opportunity for cross selling. Many banks even felt that they could not offer their retail customers an acceptable menu of banking services unless they operated a card plan. And fourth, even though achieving profitability would remain challenging, some banks had shown they could make money in the card business. The number of bank issuers grew from about 600 in 1971 to at least 1,750 in 1981.

The association model provided banks with opportunities to specialize that they didn't have with their earlier go-it-alone card programs. Banks could just issue cards, since those cards could be used at all merchants that were card customers of other members of the association. Or they could just handle credit card transactions for merchants (as acquirers). They could also do both—which is what most banks did in the 1970.

Nonetheless, echoing some of the problems of the 1950s, profits did not come easily or quickly. The early 1970s were marked by substantial losses. As described in 1971 by the *American Banker*: "The top managements of many of the nation's large credit card issuing banks, though, are becoming increasingly disillusioned with the negative profit contribution of their card programs and are questioning whether they can afford to stay in the business." Wells Fargo lost more than \$31 million between 1967 and 1970; Bankers Trust lost nearly \$27 million in 1969 and almost \$13 million in 1970; Riggs National Bank lost almost \$5 million in 1970 (as always, these figures are in 2008 dollars). All told, bank credit card losses in 1970 rose 50 percent over 1969, to \$514 million or 3.4 percent of the outstanding credit card debt.

There were a couple of reasons for the mounting losses. First, and perhaps most important, the credit card business was quite different from traditional types of lending. As one industry observer noted, banks entered the credit card industry "without being remotely prepared to solve the hundreds of problems, both large and small, that were bound to arise." Learning what worked and what didn't involved costly mistakes and inefficient practices. Second, banks did not have the luxury of starting out small while learning the business. As the *American Banker* reported in 1966: "Credit cards are not something a bank can 'feel its way into.' They require a big splash of publicity, much careful planning, aggressive selling and perhaps above all, the courage to continue as the losses mount."

The second reason behind the mounting losses of the early 1970s had its roots in the late 1960s. The "big splash of publicity" that banks used to enter the payment card industry often came in the guise of mass mailings of free, unsolicited credit cards. Banks typically drew names for the mailings from lists of depositors and customers with mortgage and installment loans. The mass mailings got the cards to lots of consumers, providing a sufficient cardholder base to attract merchants. But they also came at a price. Given the volume of consumers who received a free card in the mail and the rush to beat competitors' card offerings, the banks often did little more than a basic credit check, and some didn't even do that much. Moreover, they could not provide secure delivery of cards. Chicago postal clerks were caught hoarding unmailed cards to sell on the black market; small-time criminals stole cards out of mailboxes; even the Mafia got involved, trafficking in stolen cards and working with dishonest merchants to submit false sales slips. As a result of all this, the levels of fraud and defaults in the early 1970s were quite high. (In April 1970, the Federal Trade Commission banned the mass mailing of credit cards.) The economics of two-sided markets tells us that we should not be too quick to claim that these efforts were bad business, however. Many platforms invest in getting one or both sides on board; it is unclear whether banks could have ignited their platforms in some other way without incurring comparable losses.

Added to the turmoil of fraud and consumer delinquency were the thorny issues of how to coordinate the various banks in the system. Who was responsible for fraud and unpaid cardholder bills, and how would the various parties settle with one another? Recall that in the early 1970s, the newly formed bank associations did not have computer systems to smooth transactions; everything was done with paper and postage.

Although bankcard issuers had started to leave the red ink behind by 1972, they still found it difficult to earn the same returns from credit card lending as they did in their other lines of lending. From 1974 (the earliest year with available data) through 1980, the rate of return on assets (net before-tax earnings divided by assets) on credit card lending ranged from 1.61 percent (in 1980) to 3.09 percent (in 1977), for an average of 1.53 percent. In contrast, banks earned far higher average rates of return on other forms of lending over that same period: 2.26 percent on installment loans, 2.48 percent on real estate loans, and 3.04 percent on commercial loans. The discrepancies are especially surprising given that credit card loans are not secured with physical assets, whereas the other loan types are. Credit card loans were thus riskier to make, but earned banks a lower return throughout the 1970s.

D. And Yet, a Decade in Which Foundations Were Laid

Although the bankcard programs struggled in the 1970s, the associations to which they belonged laid the foundations for the modern electronic payment card systems. The associations established their brands—MasterCard and Visa—firmly in the minds of households and merchants. Both associations spent millions of dollars in the 1970s on national television ad campaigns, such as the 1973 "Relax—you've got a Master Charge" effort. Most important, they relied on the computer revolution to develop systems for quickly and efficiently authorizing as well as settling transactions among the growing numbers of merchants, cardholders, and members.

Visa built the BASE-I system using computers from Digital Equipment Corporation. The system allowed a merchant's authorization request to be transmitted over phone lines from the merchant to the cardholder's bank, with Visa providing backup when the cardholder's bank was closed. This went online in 1973. BASE-I cut the wait for a transaction authorization from four minutes on average to about forty seconds. It cost about $35 million to build, but it was estimated to have saved members over $110 million in fraud prevention in its first year of operation. Visa then set about building BASE-II, which computerized the entire transaction process and solved the other major member headache: the physical interchange of paper among members. BASE-II went online in 1974. MasterCard made similar investments to move its systems off paper and onto computers, with its own BankNet and I-Net systems.

V. THE 1980S' SPENDING AND DEBT SPREE

Let us take stock of the shape of the payment card industry in 1983—a year that marked the end of stagflation in the United States and the start of a long economic expansion. At least 1,500 companies ran credit or charge card programs in the United States. All of the companies with credit card programs were financial institutions that belonged to the MasterCard and/or Visa associations. Many issued cards with both brands; those that acquired transactions from merchants did so for cards from both associations. American Express was the major charge card company with 87 percent of all T&E cards; Carte Blanche and Diners Club, both operated by Citigroup, held the remainder. Table 2 lists the top issuers in 1983 of credit and charge cards by gross charge volume.

Table 2 Top issuers of credit and charge cards, 1983

Issuer	Transactions ($billions)	Share
American Express	$55.6	25.6%
Bank of America	$11.6	5.3%
Citigroup	$10.5	4.8%
Diners Club/Carte Blanche	$8.6	4.0%
First National of Chicago	$6.0	2.7%
Chase Manhattan	$5.6	2.6%
Chemical Bank	$4.2	1.9%
First Interstate Bank	$3.6	1.7%
Manufacturers Hanover	$3.4	1.6%
Security Pacific National	$3.0	1.4%
Total (top 10 issuers)	$112.0	51.6%
Total (all issuers)	$217.2	100%

Note: Numbers may not add up to totals due to rounding.
Sources: The Nilson Report; Visa U.S.A.

The payment card industry was poised for growth. Interest rate ceilings were no longer as serious a problem. Banks' cost of funds fell dramatically after the 1980–1982 recession, and usury laws were less prevalent, less severe, and after the *Marquette* decision, less important. Many banks had mature card programs with customers whose credit behavior they had observed for some time. The 1970 Fair Credit Reporting Act helped to ensure that creditors had access to accurate credit reports, increasingly at the national level. And the card industry had learned from experience how to reduce losses from cardholders who didn't pay their bills.

Credit cards were well-understood products that could make money for banks in several ways. Banks that issued cards charged membership fees. (Many banks had instituted these fees in response to a 1980 Federal Reserve requirement that they hold reserves against outstanding credit card debt.) They also earned income from merchants: they received a percentage of each transaction through the interchange fee, which was around 1.6 percent in the early 1980s. And they received finance charges from cardholders who revolved their balances. Banks that acquired transactions from merchants earned revenue from fees that were usually based on percentages of transactions volume.

The payment card industry grew dramatically over the remainder of the 1980s, during one of the longest peacetime economic expansions in U.S. history. Between 1982 and 1990, overall consumer spending rose from $4.0 trillion (about $48,000 per household) to $5.6 trillion (almost $60,000 per household). Retail spending increased from $2.1 trillion (about $25,000 per household) in 1982 to $2.7 trillion (around $29,000 per household) in 1990. Restaurant spending almost doubled from $140 billion ($1,700 per household) to $262 billion ($2,800 per household). These increases naturally provided more opportunities to use payment cards. Dollar transactions on credit and charge cards increased from $181 billion in 1982 to $467 billion in 1990. Part of this increase reflected growth in spending on cards by people who had cards. The average monthly charge on cards increased from $313 per household in 1977 to $461 in 1992. (Data are not available from the early 1980s.) Another part reflected growth in the number of households with cards. The percentage of households with at least one card increased from 43 percent in 1983 to 62 percent in 1992. (More people started carrying multiple cards so the average amount charged per card did not increase as much as the average amount charged per household.)

People not only charged more. They borrowed more. Between 1982 and 1990, the total amount of consumer credit outstanding rose from $747 billion to $1,186 billion, and the average consumer credit outstanding per household increased from almost $9,000 to just under $13,000. Thus, the total consumer credit outstanding grew at an average annual rate of 6 percent. Credit card debt grew even faster: outstanding loans on credit cards grew at an average annual rate of 21 percent, from about $58 billion to $263 billion, increasing from about $680 to slightly more than $2,800 per household in 1982 and 1990, respectively

Not surprisingly, the increased demand for payment and credit services, combined with a more favorable financial environment for payment card issuers, generated an enormous increase in the supply of payment card services. Existing issuers from Citigroup to American Express issued more cards. More important, banks that wanted to get into the card business could issue the established MasterCard and Visa brands, and consumers could use the cards they issued instantly at the millions of merchants who already took these two brands. Between 1981 and 1991, about 4,200 financial institutions became issuing members of the Visa association. (The pool of possible entrants expanded as a result of 1982 legislation that made more institutions eligible for federal deposit insurance.) Several of these entrants are particularly noteworthy, and we will discuss them shortly.

On the other side of card transactions, acquirers signed up more merchants during the 1980s—both new merchants who were part of the wave of new business formation and existing merchants who hadn't taken cards before. From 1982 to 1990, the number of U.S. merchants accepting Visa increased by 37 percent. By 1991, more than 2.5 million merchants accepted Visa cards. MasterCard experienced a similar growth in merchant acceptance.

The increased supply of card services for customers and merchants came from three other notable sources. Sears, Roebuck and Co. started the Discover Card in 1985 (and went national in 1986). Less than two years after the signature orange-and-black Discover Card was released, there were twenty-two million cards in circulation—more than Citigroup had accomplished after two decades—and $4.7 billion in receivables, ranking it third among all credit and charge cards in the United States. Over 700,000 merchants accepted the card by 1987.

Less than two years after the Discover Card launch and about twenty years after selling its credit card (the Uni-Card), American Express launched the Optima credit card. Yet what American Express thought would be a successful competitor to the bankcard issuers quickly turned into a disaster. Default rates skyrocketed and losses soared, costing the company hundreds of millions of dollars. This sophisticated company's problems make clear the complexity of the credit card business. American Express assumed that their charge card holders would be good credit risks—after all, they paid their bills on time. What American Express failed to recognize, though, was that many of these charge card holders were using their cards for company business for which they were reimbursed. They were less conscientious when they had to pay the bills themselves. There were enough defaults to sink the Optima card in red ink. By October 1991, during a recession that hit white-collar workers particularly hard, about 8 percent of Optima's receivables were charged off. The credit arm of American Express had to take a $382 million charge against third-quarter earnings in 1991, including $158 million to restructure the credit card operations and $223 million to add to credit loss reserves. In addition, about 1,700 employees were laid off. Altogether, American Express reported a 93 percent drop in its third-quarter income for 1991. By early 1992, Optima's credit losses topped out at 12 percent of receivables. American Express eventually turned its credit cards into a success in the 1990s, as we will see, but it wasn't easy.

Meanwhile, the pioneers of the payment card industry, Diners Club and Carte Blanche, had become marginal players in the United States by the end of the 1980s. Together, they had a share of less than 2.5 percent of payment card transaction volume.

In the 1980s, several significant entrants into the card business came in through MasterCard and Visa. Three giant nonfinancial firms—AT&T, General Electric, and General Motors—decided to get into the card business. MasterCard was much more enthusiastic about these nonbanks becoming part of its association than Visa was, so many of these firms issued MasterCards at first. These firms either bought banks that were members of MasterCard or Visa, or they entered into contracts with banks that were association members. Either way, it was the industrial giants whose names were prominently featured on the cards.

Many of the nonbanks' cards were tied to the business of the firm that sponsored the card. "We saw our relationships with twenty-two million calling–card customers in jeopardy," stated Paul Kahn, head of AT&T Universal Card Services Corp. Thus, AT&T linked its card to its primary business; the payment card could function as a calling card. A General Motors card allowed cardholders to accumulate points that could earn a rebate on a General Motors car. General Electric's card, however, had no such links. Instead, it offered general reward coupons that were redeemable at various retailers. The nonbank firms gained many cardholders and became some of the largest programs in the early 1990s. In fact, two of the top ten bankcard issuers in the early 1990s (by charge volume from 1991 to 1994) were nonbanks: AT&T Universal and Household Bank (which issued a General Motors card).

Some, but not all, of the nonbank card issuers were able to sustain this early growth. Still, the General Motors card is issued by Household (owned by HSBC) and enables cardholders to accumulate points toward a General Motors car. In 2009, General Electric Capital Financial was the twelfth-largest issuer, by charge volume, of general purpose cards; it issued cards for all four networks. AT&T, in contrast, sold its card portfolio to Citigroup in 1998, though the AT&T Universal card continues to function as an AT&T calling card.

Two other kinds of cards were introduced in the 1980s that were notable marketing successes. The first was proposed by one of the bank associations. In 1978, Visa introduced affinity programs that for the first time, allowed a nonmember's name or logo to be displayed on the face of the card. In 1980, however, Visa banned new affinity card programs, arguing that they tended to dilute the Visa brand, but it allowed the existing programs to continue. Five years later, faced with apparent payment card saturation, both Visa and MasterCard began allowing new affinity programs. After only one year, 296 clubs, charities, professional associations, and other nonfinancial organizations had developed Visa and MasterCard affinity programs. By 1989, there were over 2,000 affinity card programs in the United States, ranging from the Sierra Club to the National Football League to the UCLA Alumni Association. Affinity cards involving for-profit companies are also known as "cobranded" cards.

The other new type of card was introduced by Citigroup. Citigroup and American Airlines announced the AAdvantage bankcard in April 1987, marketing it to about six million members of American's AAdvantage frequent-flier program. At the time, AAdvantage was the largest-frequent flier program, and Citigroup was the largest credit card issuer in the United States with fifteen million cards issued. The card had a $58 annual fee and credit lines of up to around $58,000, which led some industry expert to assume that Citigroup was "competing for the same kinds of customers as American Express Co.'s Green and Gold Cards." By early 1993, about 1.5 million AAdvantage credit cards were in circulation.

These three payment card innovations in the 1980s—the nonbank issuers, affinity and co-branded cards, and frequent-flier rewards cards—each contributed to a dramatic rise in consumer card use over the decade. By 1991, there were almost twice as many cards per household compared to five years earlier.

VI. THE 1990S AND THE RISE OF THE DEBIT CARD

In 1990, almost 65 percent of payment cards in the United States offered revolving credit. And most households used credit cards for consumer purchases. Charge cards were used mainly for business expenses. Although ATM cards were common—accounting for about 31 percent of payment cards—and were technically debit cards (usable at those few merchants equipped to accept PIN debit), consumers hardly ever made purchases with them.

This mix of cards was quite different from that in other industrialized countries. Banks in most European countries, for instance, issued debit cards more often than credit cards. Some of these cards were deferred debit cards: the charges were accumulated and then deducted from the cardholder's checking account at the end of a monthly billing cycle. Others deducted charges right away. Credit cards were much less common. For example, virtually all cards in France and 83 percent of all cards in Germany were debit cards. In Japan, credit cards resemble Europe's deferred debit cards: consumers agree to pay the card balance at the end of the month, and issuers are authorized to directly debit a customer's account to pay the total outstanding shortly after the last day of the payment cycle.

Debit cards had been around in the United States since 1975, but they were rarely used. That changed in the 1990s as a result of two developments. After several false starts, the Visa association found a way to ignite debit cards—to get its members to issue them and households to use them. Since most of these debit cards required the cardholder to sign an authorization slip, we call them signature debit cards here. The other development came from outside the traditional payment card industry. The Electronic Fund Transfer (EFT) networks, which at the time were operated mainly as associations of banks, transformed ATM cards into debit cards during the 1990s by persuading merchants to install equipment to accept them. Since these cards required cardholders to enter their PIN, we refer to them as PIN debit cards.

The contracts that merchants entered into with Visa acquirers required them to take all Visa cards—debit cards as well as credit cards. From the standpoint of the merchant, Visa's debit cards worked just like its credit cards and didn't require any additional equipment or training. But Visa didn't have the other side of the market—the cardholders—because banks hadn't been interested in issuing these kinds of cards. As we describe in more detail in chapter 8, Visa embarked on a campaign to convince banks that they could make money from debit cards and to convince cardholders to use these cards. Visa made a commitment to promote a new debit card brand—Visa Check—through extensive national advertising. Around the same time, Visa staff tried to convince banks that the interchange fee revenues they would receive from transactions on the cards they issued would make these cards profitable. (Visa's debit card interchange fees were set slightly lower than its credit card interchange fees at the time.) MasterCard followed with a similar product and strategy after initially pursuing an unsuccessful PIN debit strategy.

The bank that operated the EFT networks faced a different business problem: they had gotten many cards into the hands of households, but didn't have merchants who were willing to take those cards. During the 1980s, banks had installed ATMs and issued cards to their customers that allowed them to take out cash and conduct other banking transactions. Associations of banks formed that enabled customers of one bank to use their cards at the ATMs of other banks. The EFT networks didn't have a system for authorizing card transactions by signature. But they could authorize and settle transactions if merchants had PIN pads that were connected to the ATM switches. And most of the EFT systems required banks to allow ATM cards to be used for retail transactions. To get merchants on board, the EFT networks had to convince merchants to install PIN pads. They did this by setting an interchange fee that was much lower than that charged by the card associations. This resulted in merchants with PIN pads paying a much lower merchant discount for PIN debit transactions than for signature debit transactions.

Most banks chose to put the marks of either MasterCard or Visa (for signature debit) and the EFT networks (for PIN debit) on the same card. This resulted in synergies between signature and PIN debit. The existing base of ATM cards made it easier for banks to issue Visa Check cards, while the Visa Check promotions helped persuade people to carry and use these cards—in both modes—for paying for things. Despite these synergies the card associations and the EFT networks waged war to capture consumer transactions. That and the competition between MasterCard and Visa for banks to issue their debit cards is one of the subjects of a later chapter on the system wars.

The efforts to increase debit card use, by both Visa and the EFT networks, were successful. By 2002, debit cards accounted for 29 percent of both all payment cards and all payment card volume. Of the $572 billion worth of transactions on debit cards in 2002, about two thirds were signature debit and one third was debit. Debit cards became an important part of the package that banks provided to consumers. In addition to providing interchange fee revenues, these cards helped banks acquire checking account customers who then typically used the bank for various services ranging from brokerage accounts to personal loans to mortgages.

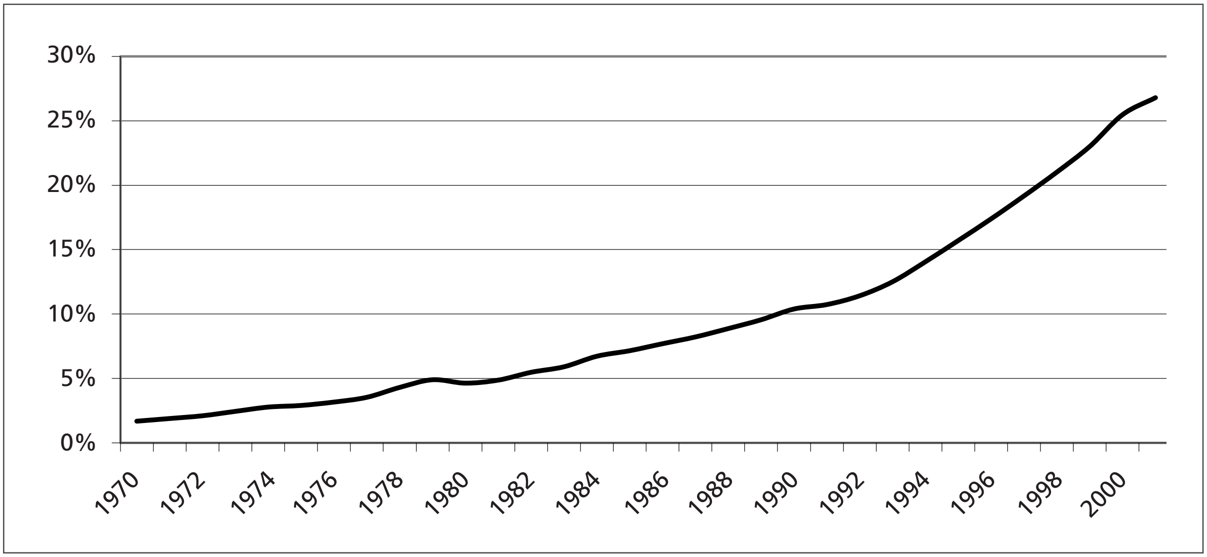
While the rapid growth of debit cards was by far the most notable change in the payment card industry in the 1990s, four other developments deserve mention. The first was a financial innovation known as "securitization," which enabled credit card lenders to sell credit card debt to other institutions that could consolidate many different kinds of debt from many different lenders. Credit card issuers could use securitization income to expand their businesses. Securitization also allowed assets to be moved from the card issuers' balance sheets, thus lowering the capital reserves they were required to hold. In addition, credit card issuers could reduce their risk from cardholders defaulting on payments—a particular worry in the event of an economic downturn. Without securitization, an individual issuer, especially those that focused on credit card lending, had difficulty diversifying away their risk exposure. Securitization allowed lenders to diversify their risk and thus to extend credit deeper into the pool of relatively risky consumers (that is, consumers with poor credit histories). Industry experts estimate that $61 billion in retail and bank credit card loans, roughly 10 percent of outstanding balances on store and bank credit cards, were securitized in 2001. (This development came later than, but is similar to, the securitization of mortgages. Most mortgage lenders resell their loans to companies that consolidate loans from many lenders and then diversify the risk through complex financial arrangements.) Securitization also made it easier for companies to both enter credit card lending (because they could diversify their risk) and get out of the business (because they had a market for the loans they had made).

The second key development occurred in the banking industry. We saw in the previous chapter that legislative changes resulted in a massive consolidation of banks. This had effects on the payment card industry. Consider the fifty-largest banks ranked by total assets in 1990. Through mergers and acquisitions, these fifty banks were consolidated into eighteen banks by 2003. For example, Chase Manhattan took over Chemical Bank in 1996. Bank One acquired First USA and First Chicago in 1997 and 1998, respectively, and Bank of America—one of the largest bank credit card issuers—merged with NationsBank in 1998. This automatically led to consolidation among payment card issuers. Moreover, there was substantial consolidation within the card industry through portfolio sales. AT&T Universal sold off its portfolio to Citigroup in 1998, for instance, and MBNA purchased the SunTrust Bank portfolio in 1999. Additional shifts have occurred as some issuers have substantially increased their share of sales volume. For example, MBNA more than doubled its share of Visa and MasterCard credit card volume from 4.5 percent in 1990 to 10.8 percent in 1999. The payment card industry became more concentrated—the top ten bankcard issuers accounted for 74 percent of Visa and MasterCard volume in 1999 versus 44 percent in 1990.

A third development concerns the co-opetitives. During the 1990s, each co-opetitive began aggressively encouraging its members to dedicate themselves to one association at the expense of the other association. By the end of the 1990s, both MasterCard and Visa began to offer formal partnership programs, providing benefits to banks that agreed to focus their card business on just one system.

Finally, one can't talk about the 1990s without mentioning about the emergence of the commercial world-wide web in the middle of the decade. Cards were the perfect payment instrument for online transactions and quickly became the main way to pay on the web.

Figure 4 Purchase volume on general-purpose payment cards as a percent of consumer expenditures, 1970-2000



Note: General-purpose payment cards include credit, charge, PIN debit, and signature debit cards.

Sources: Bureau of Economic Analysis; and *The Nilson Report* (various 1970–2003 issues).

VII. GOLDEN ANNIVERSARY

The general-purpose payment card turned fifty in 2000. Despite all the talk about the cashless society, this electronic method of payment still accounted for less than half of all retail transactions, and only 27 percent of all consumer expenditures—this excludes implicit consumption that doesn't involve payment, such as the rent you pay yourself if you own your home. Cash and checks still hadn't disappeared.

The move to electronic money had been gradual, but steady during the first half century. Figure 4 shows the purchase volume of general-purpose payment cards as a percentage of consumer expenditures in the United States from 1970 to 2000. (Comparable data are not available before 1970. The percentages would have been miniscule during the 1950s and through much of the 1960s.) The payment card industry had evolved through incremental changes that persuaded more individuals and businesses to rely on this method of payment. There had been some drastic innovations—McNamara's initial insight, and the idea of bundling revolving credit with a payment device, and the ignition of debit. But modest innovations had also been important in aggregate. These include the use of computers to reduce the time it takes to complete transactions, the development of credit-scoring techniques to identify and then monitor creditworthy customers, securitization, and bundling airline rewards and other features onto cards.

VIII. NEW BEGINNINGS

Standing at the end of the 1990s one might have thought that the payment card industry, at fifty, had reached maturity, that its sixth decade would be relatively calm. It was not to be. By 2010 the MasterCard and Visa associations had dissolved themselves and reemerged as publicly traded companies with market caps, when combined, of around $85 billion—three times larger than eBay and almost 60 percent as large as Google.[1] Credit cards showed their age as debit cards soared and prepaid cards lay the basis for considerable innovation in payments. The industry ended the decade under siege as Congress, responding to a barrage of complaints from consumer and merchant groups and profoundly cynical about financial service providers, imposed intrusive regulations. Yet, while old ways were under challenge, innovation accelerated as a result of developments with mobile, software and internet technologies. Many industry participants ended the 2000s with fright and delight as they dealt simultaneously with significant business risks and substantial new opportunities.

A. Death by Association

By the turn of the century the bank associations that had spurred the growth of cards through network effects for three decades were facing serious business and legal issues.

The credit-card associations were clubs that bank members belonged to. Like any club management spent a far amount of time making sure that the members were on board with their various initiatives. And, like any club, there was an inherent tension between management, which had to set and enforce membership rules, and the members who could vote management out. These clubs became harder to run during the 1980s and 1990s as a result of increasing tension between the bank members and between members and management.

In the late 1960s and 1970s credit-card issuers were fishing for customers in a very large ocean. Before the Marquette decision opened the way for national card issuance many of these banks were also fishing far away from each other. They issued nationally and competed at least in part for the same customers. Securing cooperation among these rivals became harder for the managements of MasterCard and Visa.

At the same time the larger banks became better able to seek advantages for themselves from each association by threatening to take their card portfolios to the other association. That wasn't possible in the earlier days. Even the withdrawal of the largest members of MasterCard and Visa wouldn't have had a noticeable effect on the financials of these associations in the 1970s and 1980s. While many banks belonged to both associations they could threaten to move their volume over the other association if they didn't get what they wanted in return.

[1] Market caps as of July 17, 2009.

Citigroup did just that. Several banks including Citi wanted Visa to give the bank brand more prominence on the card. Up to that point in time MasterCard and Visa cards highlighted the brands of the association. That made sense during the years in which all banks benefit from the expansion of these card brands in the United States and, indeed, around the world. But as the banks became bigger and more ambitious rivals, and the MasterCard and Visa brand names were engraved in the minds of the public, some of the banks wanted the ability to give themselves a larger piece of the real-estate on the plastic card. Despite Citi's threat to bolt Visa refused to relent to Citi's demand over this or other issues. This national issuer, made good on its threat and started that year to move its U.S. card volume over to MasterCard. That was a financial jolt to Visa and to its members who had to bear a greater part of the cost of the system.

Ironically, it is easier to manage a club with many small members than one which a handful of members who are trying to seek advantages for themselves. These "clubs" had that worked well for many decades became a victim of their own success as some their members became very large.

The associations had been flypaper for litigation almost from their beginnings but the risks had become both more palpable and larger in the early days of the new century. Even without the management challenges MasterCard and Visa had little choice but to kill the associations and reincarnate themselves as public equity corporations.

The antitrust laws of the United States and many other countries are rightfully suspicious when firms that compete with each other get together and agree to do things. The most severe antirust sanctions are imposed on "cartels" in which competitors get together to fix prices or divide markets. Yet the U.S. Supreme Court has recognized that under certain circumstances cooperation among competitors may generate value for consumers. The European Union's antitrust laws also allow firms to cooperate when they can show, in effect, that this is essential for providing benefits to consumers. Over the years the courts and regulatory authorities have struggled with deciding whether these associations of banks rivals are restricting competition or engaging in necessary cooperation that promotes economic efficiency. But any action that MasterCard and Visa take can be subject to a lawsuit, or a complaint to an antitrust authority, that the banks are violating antitrust laws that prohibit competitors from acting together. The burden then falls to the associations to prove that they should get a pass.

The greatest controversy over the years has surrounded the interchange fee. MasterCard and Visa set the fees that members that acquire merchants have to pay members the acquire cardholders when a cardholder uses her card to buy something at a merchant. Antitrust and bank regulators around the world started looking at interchange fees in the late 1990s. Then, in 2005, a group of retailers in the United States, including many large ones, filed class-action action cases against MasterCard and Visa accusing them of having engaged in price fixing and other violations of the antitrust laws. It became apparent early in the decade that the associations were going to be embroiled in litigation and regulatory proceedings over the setting of interchange fees for years, that interchange fees in various part of the world could be prohibited or regulated, and that in the United States at least associations (and their members) could face massive damage awards. MasterCard and Visa agreed to pay about $3.3 billion in 2003 (in 2008 dollars) to settle another merchant class action. The merchants had claimed that the card networks have restricted competition by "tying" credit and debit cards—merchants had to take both to get either.

The legal issue at the heart of these lawsuits results almost entirely from the association model that MasterCard and Visa adopted. The interchange fee largely determines the fee that merchants pay because it is passed on from the acquiring bank to the merchant. Go-it-alone systems such as American Express can set the merchant fee for their cards. They are just setting the price they are charging customers. They can also set the portion of the fee that they share with banks that agree to issue American Express cards. MasterCard and Visa were vulnerable to lawsuits largely because members that competed with each other were acting through the association to set the interchange fee.

The banks that owned MasterCard and Visa decided to end their associations. MasterCard went first in 2006 and Visa next in 2008. In both cases these card networks reorganized themselves as public equity corporations and did initial public offerings in which a majority of the stock was sold off. Banks retained some stock, continued to have representation on the boards, but removed themselves from any decisions concerning interchange fees. These IPOs allowed the banks to cash-in on their participation in the card associations. By allowing these organizations to raise capital in the public markets, and eliminating the cumbersome association management, the banks likely increased the value of MasterCard and Visa and thus their own stakes. And finally, by removing the banks from a decision making role over interchange fees the banks arguably reduced their exposure to litigation involving price fixing claims. (The plaintiffs in the US class-action litigation and the European antitrust regulators, however, continue to claim that the banks are colluding unlawfully in the setting of interchange fees.)

Another significant change in the role of card networks occurred in the 2000s as a result of a lawsuit that had been filed by the U.S. Department of Justice. The courts ruled that it was unlawful for MasterCard and Visa to prohibit member banks from also issuing the cards of competing associations. (American Express had already been issuing cards through MasterCard and Visa member banks outside the United States where the association gave in to the threat of litigation.) American Express and Discover both have "network" divisions which enter into agreements with banks to issue their cards. For the United States, these usually go-it-alone systems have not reported how much of their card transactions occur on cards issued by bank partners although there is no evidence that it is significant as of 2010. (Outside the US American Express reports that 68 percent of its transactions occur on cards issued by third-parties on its network.)[2]

[2] American Express 2009 10k.

B. Pay Before and Pay Now

For its first half century the defining characteristic of the American payment card industry for consumers was that when they used a card they didn't have to really pay until later. The tide turned in the 2000s.

Debit cards use grew explosively. By the middle of the decade, debit cards accounted for more than half percent of all general purpose card transactions and more than a third of all dollars charged to cards.[3] For Visa, the first quarter of 2009 marked a significant turning point: debit card charges exceeded credit card charges for the first time.[4]

Credit cards allowed people to pay later by writing a check for the funds they were advanced today to buy something. Debit cards enabled people to pay more or less right away by having funding taken out of their checking accounts. The timeline was completed by a product that was developed in the mid 1990s but became widely developed in the 2000s: the prepaid card that allowed people to commit funds to a card and have purchased deducted from that account. The concept of "pay before" cards became the mother of invention in the card industry: it spawned some of the most important innovations in payments, from enabling the widespread disbursement of funds to the largely unbanked victims of Hurricane Katrina, to providing a low-cost way for companies to pay their employees, to helping immigrants to get funds cheaply to their families back home.

American retailers started replacing paper gift certificates with plastic gift cards in the mid 1990s. Blockbuster was the first in 1995. A gift giver could buy a gift card that had a designated amount of funds attached to it. The recipient of the gift card could then use that card to buy items at the retailer until they had exhausted the funds on the card. Other retailers such as Starbucks introduced cards that could be reloaded with money. People could buy the cards for themselves, or give them away, and pay for items at the retailer. (Retailer prepaid cards are often called "closed-loop" cards there is a direct flow of funds between the merchant that issued the card and the consumer who uses the card.)

General purpose prepaid cards were introduced at the turn of the century by the major card networks. Once loaded, cardholders could dip into the funds for these cards at any merchant that accepted the card brand. Moreover, some of these general purpose prepaid cards were linked with a PIN-debit network so that consumers could withdraw funds at almost any ATM. Many uses were found for these prepaid cards and they are often referred by their particular use.

Issuers persuaded some employers—sometimes working through the payroll processing firms that employers use—to give their employees "payroll cards" and make their wages accessible

---

[3] According to data from the Federal Reserve's 2007 Electronic Payments Study, signature and pin debit cards accounted for 57% of all general purpose card transactions in 2006 and 35 percent of all general purpose card volume (this includes excludes store cards and prepaid cards). About two-thirds of both transactions and volume go through MasterCard or Visa with a signature and one third through one of the PIN-debit networks.]

[4] Robin Sidel of the Wall Street Journal

through these cards. The employer would use the ACH system to deposit wages into an account at the prepaid card issuer. The employee could then use their card to either pay for things at merchants that accept the card brand or to take money out of ATMs. Payroll cards are particularly helpful for employees that don't have bank accounts.

Prepaid cards were also appealing for the unbanked more generally. Aside from convenience, by the 2000s it had become difficult for a consumer to carry on the ordinary business of life without a plastic card. They were essential for buying things online, a few places had stopped accepting cash, and some transactions required a card for a security deposit. The unbanked overlapped significantly with another group of people who bound prepaid cards useful: immigrants who wanted to transfer funds to family in friends in other countries. They could do this through money transfer services such as Western Union but these services, which rely on labor-intensive networks of offices throughout the world, were expensive. Prepaid card issuers introduced cheaper card-based alternatives.

Prepaid card solutions were also developed for governments and non-profits. Many states use prepaid cards to distribute benefits such as food stamps. Visa alone ran 70 state prepaid programs in 38 states in 2009. They load child support, workers comp and unemployment insurance payments on these cards. The American Red Cross provided massive assistance to the Hurricane Katrina victims on prepaid debit cards from MasterCard and JPMorgan Chase and is using these cards as part of its general relief efforts.

Although there is a consensus across many data sources that prepaid cards have grown over the 2000s there are no reliable estimates of the amount of spending that takes place on these cards. (That is because the main data collection schemes cannot readily distinguish between prepaid and debit cards.) A detailed study sponsored by the Board of Governors of the Federal Reserve estimated that, in 2006, consumers charged $58.3 billion on retailer and general purpose prepaid cards.[5] Most of the prepaid spending, 73.3 percent was on retailer cards. Prepaid cards accounted for a small portion of overall payment card spending mid decade. Retailer and general purpose prepaid cards amounted to 1.6 percent of total spending on all retailer and general purpose payment cards in 2006. General purpose prepaid cards accounted for only 0.5 percent of spending on general purpose payment cards.

C. The Financial Crisis and the New Rules

The decade ended with Congress imposing extensive regulations on the payment card industry that that could have far-reaching consequences on the business in the coming years.

Cardholders complained about issuers charging then hefty late fees while giving them little time to pay their bills. Issuers were getting more of their revenues from these soft of fees.

[5] Federal Reserve Board, Electronic Payments Study, 2007.

Consumers were also irritated by banks increasing interest rates on their cards without notifying them and applying these higher rates to existing balances. The press highlighted extreme cases of bad behavior by issuers and consumer advocates agitated for relief.

Meanwhile, merchants, the other half of the two-sided payments business, became increasingly vociferous about the fees they had to pay when consumers paid with plastic. Their bills had increased over the decade in part because interchange fees had increased and because consumers were paying more with cards (in particular with debit cards). In addition to pursuing an antitrust lawsuit they lobbied Washington for relief.

We will never know whether Congress would have adopted legislation to deal with either of these issues if the financial crisis hadn't hit in 2008. The banking industry was blamed for having caused the Great Recession and forcing taxpayers to fund massive bailouts to stem financial collapse. Calls for financial reform soon followed. Although the crisis was largely caused by the burst of a housing bubble the payment card industry was swept along in efforts to devise stricter regulations for banks and curtail what some thought as excessive credit.

The Credit Card Accountability and Disclosure Act (CARD Act), signed by President Obama in May 2009 after having won bipartisan support in Congress, regulated many aspects of how issuers could charge consumers. The Act prohibited some practices such as raising interest rates in the first year after a consumer has opened a credit card account, charging for exceeding limits, and setting early morning deadlines for payments. It also required issuers to give consumers at least 45 days notice of increase in interest rates and to mail bills at least 21 days in advance of the due date of the payment. Credit card companies complained that the legislation limited their ability to charge more to riskier borrowers. Although systematic data on the results of the legislation are not yet available there were reports that banks raised card fees and interest rates to cover their lending risks and reduced lending to riskier borrowers. They were able to respond that way because the CARD Act did not limit their fees—only the timing of them.

Slightly more than a year later Congress, divided along party lines, passed the Wall Street Reform and Consumer Protection Act which was signed by President Obama on July 21, 2010. Two parts of this legislation, which was the major effort to reform financial regulation in light of the crisis, affect the payment card industry.

The most direct is a section which calls for "reasonable fees and rules for payment card transactions." The Federal Reserve Board is supposed to make sure the debit card interchange fees are reasonable and proportional to the costs incurred in making the transaction. That applies to signature and PIN debit as well as most general purpose prepaid cards. Since banks typically earn most of their debit card revenue from merchants rather than cardholders these rules will likely reduce interchange fees, especially for signature debit, substantially. The Federal Reserve is supposed to have its regulations in place by the end of March 2011.

The legislation also creates the Consumer Financial Protection Board (CFPB), within the Federal Reserve Board, which is responsible for regulating almost all consumer financial service products including all payment cards. In part, the CFPB takes over existing responsibilities under past legislation from various federal agencies. But it also has new powers to prevent financial service companies from engaging in "abusive" practices. The legislation furthermore allows states to adopt and enforce more stringent rules and laws than those adopted by the federal government. It remains to be seen how this new agency will affect the card industry. It will take time for it to be established and much of its powers are vested in its director who will be appointed by the President and confirmed by Congress.

D. Inflection Point

Shortly after the start of the financial crisis eBay bought BillMeLater for almost $1 billion in cash and stock with plans to add it to its PayPal business. BillMeLater provided credit to consumers for online transactions. Consumers who shopped at merchants that offered this alternative could seek an almost instant credit approval for their purchase. That might involve a revolving loan or an installment purchase. If they got it they got the good and got the bill later. In late 2009 American Express spent $300 million to buy Revolution Money which used Web 2.0 technology to provide cards online and as well as in the physical world. Less than two years later Visa bought Cybersource, a company that provided various technical services to online merchants, for about $2 billion. These transactions were proof of what was widely known in the payments industry. These transactions sent a message to entrepreneurs and investors in the payments business: innovation could have big payoffs.

The business of using cards for buying and borrowing was ripe for new ways of doing things at the end of the decade. Most importantly, the walls between the online and physical worlds were falling. For most of the web economy's short history there was a sharp distinction between shopping online, which people did sitting at their desktop computers using a browser, and at bricks and mortar stores that people drove or walk to and where they could touch what they were thinking about buying.

By 2009 about 91 percent of American households had mobile phones[6] and about 38 percent of subscribers used them to access the internet.[7] The percent of consumers with smart mobile phones—basically small computers with a graphical user interface and a browser—such as the iPhone was increasing rapidly. A few of the new payment services pointed toward future possibilities with the mobile devices. People could use PayPal—the quintessential online payment method—to rent a lawnmower (through a new company called Rentalics that relied on PayPal for payments) or to split the cost of a meal at a restaurant (through the iPhone bump application) with a few taps on their smart phones.

---

[6] http://www.ctia.org/advocacy/research/index.cfm/AID/10323

[7] http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Mobile_Access_2010.pdf

Merchants also started seeing the possibility of using internet technology to accept payments at the physical point of sale. They had relied on private networks and devices that came bundled with software. The company Square introduced a new payment system that enabled merchants to accept payments using internet-connected mobile devices such as the iPad or Android phones. They provided an attachment to these devices through which merchants could swipe cards. The transaction was then sent to the merchant processor over the internet.

For most of the previous century Americans used some form of "card", whether metal, paper, cardboard, or plastic, to identify themselves to merchants. The mag stripe plastic card was the predominant method for paying at merchants in 2010—despite a well-funded effort by the card networks to persuade merchants to install equipment to take contactless cards that consumers could wave at terminal. Swipe transactions took just a few seconds to complete. Yet many were betting that the mobile phone would topple the card. They just weren't sure how or when.

As cards entered roughly their second century and the general purpose payment card turned sixty, paper was still not vanquished. Almost half of consumer expenditures—even excluding mortgages and rent—were paid with cash or check or some other paper method of payment. That means that the electronic payments industry, and the general purpose card portion of it, continues to have significant opportunities for growth.

CHAPTER

12 Interchange Fees and Their Regulation

David S. Evans and Richard Schmalensee

ABSTRACT:

This chapter surveys the economic literature on interchange fees and the debate over whether interchange should be regulated and, if so, how. We consider, first, the operation of unitary payment systems, like American Express, in the context of the recent economic literature on two-sided markets, in which businesses cater to two interdependent groups of customers. The main focus is on the determination of price structure.

We then discuss the basic economics of multi-party payment systems and the role of interchange in the operation of such systems under some standard, though unrealistic, simplifying assumptions. The key point of this discussion is that the interchange fee is not an ordinary price; its most direct effect is on price structure, not price level. We then examine the implications for privately determined interchange fees of some of the relevant market imperfections that have been discussed in the economic literature. While some studies suggest that privately determined interchange fees are inefficiently high, others point to fees being inefficiently low. Moreover, there is a consensus among economists that, as a matter of theory, it is not possible to arrive, except by happenstance, at the socially optimal interchange fee through any regulatory system that considers only costs. This distinguishes the market imperfections at issue here for multi-party systems from the more familiar area of public utility regulation, where setting price equal to marginal cost is theoretically ideal.

Next, we consider the issues facing policy makers. Since there is so much uncertainty about the relation between privately and socially optimal interchange fees, the outcome of a policy debate can depend critically on who bears the burden of proof under whatever set of institutions and laws the deliberation takes place. There is no apparent basis in today's economics - at a theoretical or empirical level - for concluding that it is generally possible to improve social welfare by a noticeable reduction in privately set interchange fees. Thus, if antitrust or other regulators had to show that such intervention would improve welfare, they could not do so. This, again, is quite unlike public utility regulation or many areas of antitrust including, in particular, ordinary cartels. By the same token, there is no basis in economics for concluding that the privately set interchange fee is just right. Thus, if card associations had to bear the burden of proof -for example, to obtain a comfort or clearance letter from authorities for engaging in presumptively illegal coordinated behavior - it would be difficult for them to demonstrate that they set socially optimal fees.

We take a pragmatic approach by suggesting two fact-based inquiries that we believe policymakers should undertake before intervening to affect interchange. First, policymakers should establish that there is a significant market failure that needs to be addressed. Second, policymakers should establish that it is possible to correct a serious market imperfection, assuming one exists, by whatever intervention they are considering (such as cost-based regulation of interchange fee levels) and thereby to increase social welfare significantly after taking into account other distortions that the intervention may create. We illustrate both of these points by examining the recent Australian experience.

I. INTRODUCTION

In 1958 the Bank of America began operating the BankAmericard credit card system, the predecessor of Visa, as a *unitary* system.[1] It performed both the issuing function (dealing with cardholders) and the acquiring function (dealing with merchants) itself. Similarly, it set the fees charged to both these customer classes—the annual fee, interest rate, late fees, and other fees charged to cardholders; and the per-transaction fee to merchants known as the merchant discount. It was therefore able to determine both the overall level of fees (which might be measured as total fees per dollar of transactions) and their structure (which might be measured by the shares of total fees paid by merchants and cardholders).

In 1966 the Bank of America began to bring other banks into the system as franchisees. Individual banks within the system were free then, as now, to determine the fees they charged merchants and cardholders. When a consumer holding a card issued by bank A made a purchase at a merchant that had bank A as its acquirer, bank A could, if it wished, have the same fee structure as Bank of America. But what if this same consumer made a purchase from a merchant acquired by bank B? Bank of America required the acquiring bank to pass the full merchant discount to the issuing bank. Acquiring banks had incentives to lie about their merchant discounts under this rule, as issuing banks were well aware. More importantly, this rule meant that acquiring banks received zero revenue for transactions for which they provided the merchant but had not issued the card being used. The rule therefore blunted the incentives for all banks to sign up merchants, to the obvious detriment of the system as a whole.

In 1970 the BankAmericard system was converted into a membership corporation, a *multi-party* system. This cooperative association established an *interchange fee* in 1971 to deal with transactions in which issuing and acquiring banks were different. This fee was paid by the acquiring bank to the issuing bank and initially set at 1.95 percent. It was not linked to any individual bank's

[1] This discussion in this paragraph and the next two follows Evans and Schmalensee 2005, pp. 153-156.

merchant discount. The interchange fee thus became a revenue source on the issuing side of the credit card business and a cost element on the acquiring side. Acquiring banks had to charge a merchant discount that was greater than the interchange fee to recover this cost. The interchange fee was an element of a standard contract that the multi-party system established for its members; other terms of the contract defined who bore the risk of fraud or nonpayment, as well as how disputes would be resolved.

This chapter surveys the economic literature on interchange fees and the debate over whether interchange should be regulated and, if so, how.

A. What's Interesting About Interchange?

Until 1979, few outside the Visa and MasterCard systems had any idea what an interchange fee was. In that year, the National Bancard Corp. (*NaBanco*) filed a lawsuit contending that when the Visa member banks determined the interchange fee, those banks engaged in illegal price-fixing, which, it was claimed, had damaged *NaBanco*.[2] An appeals court found in favor of Visa in 1986, holding that the interchange fee had potential efficiency benefits for a two-sided system:

> Another justification for evaluating the [interchange fee] under the rule of reason is because it is a potentially efficiency creating agreement among members of a joint enterprise. There are two possible sources of revenue in the VISA system: the cardholders and the merchants. As a practical matter, the card-issuing and merchant-signing members have a mutually dependent relationship. If the revenue produced by the cardholders is insufficient to cover the card-issuers' costs, the service will be cut back or eliminated. The result would be a decline in card use and a concomitant reduction in merchant-signing banks' revenues. In short, the cardholder cannot use his card unless the merchant accepts it and the merchant cannot accept the card unless the cardholder uses one. Hence, the [interchange fee] accompanies "the coordination of other productive or distributive efforts of the parties" that is "capable of increasing the integration's efficiency and no broader than required for that purpose."[3]

In 1983, William Baxter, a leading antitrust scholar who had worked for Visa on the case, published an important paper on the economic rationale for interchange fees in 1983.[4] But after the *NaBanco* decision, interchange fees faded from view in academic and policy circles and was a topic of interest mainly to industry insiders.

A few academic papers in the 1990s mentioned interchange fees,[5] but for the most part, this topic languished in obscurity until around the turn of this century. And then, as the dates on

[2] National Bancard Corporation v. Visa U.S.A., Inc. 1984, 1986a, and 1986b.
[3] National Bancard Corp. v. Visa U.S.A., Inc. 1986a, p. 602.
[4] Baxter 1983. We discuss this paper below.
[5] These include Carlton and Frankel 1995a and 1995b, Evans and Schmalensee 1995, and Frankel 1998.

most of the entries in this chapter's list of references indicate, interest in interchange fees increased dramatically among academics, banking regulators and competition authorities around the world.

Two developments caught the attention of policymakers. Cards had become an increasingly important part of the payment system in many countries. The share of consumer expenditures in the United States paid for with cards had increased from about 3 percent in 1986, the year of the *NaBanco* decision, to 25 percent in 2000.[6] Similar increases occurred in other countries. For example, in 2000, cards accounted for 30 percent of consumer expenditures in Australia and 35 percent in the United Kingdom.[7] While American Express played a significant role in the United States, globally most cards were associated with multi-party systems that had interchange fees. In the United Kingdom, for example, an influential report on the banking industry issued in 2000—the "Cruickshank Report"—addressed interchange fees in multi-party systems, concluding that "[t]here is a strong case for reform of the interchange fee system."[8]

For retailers, merchant discounts (which included interchange fees in the bank card systems) had become a growing portion of their costs, as more people paid with cards. Some retailers had periodically complained about merchant discounts; hotels had gone so far as to create their own card system in the United States in the mid 1950s to avoid the merchant discount of Diners Club (a unitary system). However, with increases in interchange fees and perhaps other legal and political developments, various organizations of retailers around the world sought regulatory relief from the fees. For example, EuroCommerce, a retailer association, filed a complaint with the European Commission in 1997. This led to an investigation of Visa Europe and ultimately a settlement in which Visa Europe agreed to lower the interchange fees. Interchange fee levels have also been under active attack and/or regulated in a number of other countries, including Australia and the United Kingdom.[9]

Why has interest in interchange fees surged among economists in recent years? Some, including us, were exposed to the topic through their involvement in litigation and/or regulatory proceedings. But the same could be said for many other legal and regulatory issues that have spawned much smaller literatures over longer periods of time. The large volume of theoretical literature on interchange fees has arisen for the simplest of reasons: understanding their determination and effect is intellectually challenging. As the discussion below indicates, this is not necessarily good news for policy-makers.

---

[6] HSN Consultants 1987, 1991, 1992, 2001a, and 2001b, and United States Department of Commerce2003.

[7] Visa International 2001. Measures of consumer expenditures may not be entirely consistent across countries but are sufficiently comparable to provide a rough sense of importance. See also Weiner and Wright 2005.

[8] Cruickshank 2000, p. 272.

[9] See Reserve Bank of Australia 2002, European Commission 2002 and 2003, and Office of Fair Trading 2003. Weiner and Wright (2005) provide a useful overview of government involvement.

B. This Chapter

The remainder of this chapter is organized as follows. The next section provides context by considering the operation of unitary payment systems, like American Express, in the context of the recent economic literature on two-sided markets, in which businesses cater to two interdependent groups of customers. The main focus is on the determination of price structure. We then discuss the basic economics of multi-party payment systems and the role of interchange in the operation of such systems under some standard, though unrealistic, simplifying assumptions. The key point of this discussion is that the interchange fee is not an ordinary price; its most direct effect is on price structure, not price level. While it is clear that an unregulated monopolist or a cartel in a one-sided market, like electricity generation, would set a price that is higher than would be socially optimal, no such presumption exists for the interchange fee set by a monopoly card system.

We then consider implications for privately determined interchange fees of some of the relevant market imperfections that have been discussed in the economic literature. While some studies suggest that privately determined interchange fees are inefficiently high, others point to fees being inefficiently low. Moreover, there is a consensus among economists that, as a matter of theory, it is not possible to arrive, except by happenstance, at the socially optimal interchange fee through any regulatory system that considers only costs. This distinguishes the market imperfections at issue here for multi-party systems from the more familiar area of public utility regulation, where setting price equal to marginal cost is theoretically ideal.

The penultimate section examines the implications of the results of the previous sections for policy makers. Since there is so much uncertainty about the relation between privately and socially optimal interchange fees, the outcome of a policy debate can depend critically on who bears the burden of proof under whatever set of institutions and laws the deliberation takes place. There is no apparent basis in today's economics—at a theoretical or empirical level—for concluding that it is generally possible to improve social welfare by a noticeable reduction in privately set interchange fees. Thus, if antitrust or other regulators had to show that such intervention would improve welfare, they could not do so. This, again, is quite unlike public utility regulation or many areas of antitrust including, in particular, ordinary cartels. By the same token, there is no basis in economics for concluding that the privately set interchange fee is just right. Thus, if card associations had to bear the burden of proof—for example, to obtain a comfort or clearance letter from authorities for engaging in presumptively illegal coordinated behavior—it would be difficult for them to demonstrate that they set socially optimal fees.

We take a pragmatic approach by suggesting two fact-based inquiries that we believe policymakers should undertake before intervening to affect interchange. These inquiries are premised on the view, which we believe is now widely held, that the government should intervene in markets only when there is a sound basis for believing that it can devise policies that will improve social welfare significantly. First, policymakers should establish that there is a significant market failure that needs to be addressed. To do so they would need to examine the marginal social benefits and costs of alternative payment systems as they vary among transactions; we suggest that there is highly incomplete information available on these benefits and costs, so that any inference from the current data is at best problematic. Second, policymakers should establish that it is possible to correct a serious market imperfection, assuming one exists, by whatever intervention they are considering (such as cost-based regulation of interchange fee levels) and thereby to increase social welfare significantly after taking into account other distortions that the intervention may create. We illustrate both of these points by examining the recent Australian experience.

The final section summarizes our conclusions. Many of the results of the economic literature necessarily depend on various simplifying assumptions. In this last section we highlight those that we believe are robust, in the sense that they are likely to hold generally. Our main focus throughout is on conceptual issues; the companion paper by Weiner and Wright provides a good deal of useful factual material.[10]

II. PAYMENT SYSTEMS AS TWO-SIDED PLATFORMS

In the last few years, economists have come to understand that payment systems have much in common with auction houses, exchanges, shopping malls, and video game consoles. All are examples of two-sided (or, more generally, multi-sided) platform businesses. Such businesses are intermediaries that add value if and only if they can appropriately coordinate the demands of two distinct groups of customers. Beauty salons may attract both men and women, for instance, but heterosexual singles bars *must* attract both men and women—and in the right proportions. Similarly, shopping malls must attract both retailers and shoppers, auction houses need both buyers and sellers to stay in business, sellers of video games need both game players and game creators—and payment systems need both consumers and merchants.

The earliest use of the term "two-sided" in this sense of which we are aware is in a 1998 paper considering a match-making intermediary that adds value by bringing individuals of two different types together.[11] The authors find that, under plausible conditions, one of the types will not pay for the service, and they note that this highly asymmetric pricing is descriptive of some real two-sided markets, such as real estate agents in the United States and some dating services.

The reason why skewed pricing can happen is central to the analysis of two-sided markets in general: there are typically positive indirect network externalities between the two groups.[12] For example, the more men (women) who use a particular dating service, the more attractive the service is to women (men). If the service cannot attract both in the right proportions and sufficient

[10] Weiner and Wright 2005.

[11] Van Raalte and Webers 1998. While he used different terminology, Baxter 1983 did the first two-sided analysis, as we discuss below.

[12] See Ackerberg and Gowrisankaran 2003 for an attempt to quantify these externalities in a payments context.

numbers, it will fail. If it needs to serve men for free to accomplish this, it is rational for it to do so. Giving the service away to men, even though serving them involves positive marginal costs, can be profitable if it attracts sufficient fee-paying women.

The recent general literature on two-sided platform markets began around 2002 with early versions of a seminal contribution by Jean-Charles Rochet and Jean Tirole.[13] Early versions of an important related paper by Mark Armstrong appeared shortly thereafter.[14] These papers pointed out that many businesses or markets can usefully be thought of as two-sided, and they developed some general implications of the importance of balanced participation from the involved customer groups (or, as it is more commonly put, "getting both sides on board"). Since then the literature on two-sided platform markets has grown explosively.[15] This literature generally considers one or more vendors dealing directly with both (or all) involved customer groups and it is thus most directly relevant in our context to the analysis of unitary payment systems.

A. Pricing by Two-Sided Platforms

Almost all theoretical analyses of pricing by two-sided platforms assume that they either charge only an access or membership fee (following Armstrong) or that they charge only a variable or per-transaction fee (following Rochet and Tirole). While this simplification generally facilitates understanding of basic principles, it is problematic in some cases. Newspapers, for example, typically charge an access fee to readers (a fixed cost for the newspaper regardless of what is read), while advertisers pay a variable fee based on the number of readers. In the payment card context, for instance, merchants incur small fixed access (terminal) costs of card acceptance as well as per transaction merchant discounts. On the other side of that market, consumers sometimes pay membership fees, but variable fees are typically slightly negative as a result of free float and sometimes noticeably negative as a result of transaction-based reward programs.[16]

The literature on pricing by two-sided platforms distinguishes between the level of prices and the structure of prices. The profit-maximizing price level—the price paid by men plus the price paid by women in the dating service example, for instance—depends on the costs of serving both groups, on the sensitivities of both demands to price, and on the indirect network effects between the two customer groups. The price sensitivities depend in the usual way on the

[13] Rochet and Tirole 2003b.
[14] Armstrong 2004.
[15] Contributions that look beyond payment systems include Armstrong 2005, Armstrong and Wright 2004, Bolt and Tieman 2004a and 2004b, Caillaud and Jullien 2003, Gabszewicz and Wauthy 2004, Hagiu 2004a and 2004b, Jullien 2004, Rochet and Tirole 2004a and 2004b, and Schiff 2003. Evans 2003a and 2003b and Wright 2004b provide accessible discussions of the implications of some of this work as well as empirical support; Rochet and Tirole 2004b provide a more technical overview.
[16] As we note below, the payment card industry has followed this pricing structure since its birth, as a two-sided platform, in 1950.

price and quality of available substitutes and other factors, including, in some models, the presence of two-sided competitors.

The profit-maximizing *price structure*—the ratio of the price paid by men to the price paid by women, for instance—also depends in general on costs, price sensitivities, and the way participation by the members of each group affects the demand of the other group.[17] In addition, all general models of two-sided platform markets imply that profits may be maximized by highly asymmetric pricing in which one group is served at a price close to or even below marginal cost, and most or all gross margin is earned by serving the other group.[18]

It is important to note that many, if not most two-sided markets exhibit this sort of asymmetry in pricing and gross margin generation.[19] Shopping malls, for instance, often provide free parking to consumers, sometimes in expensive parking structures, and make all their money by charging rent to merchants. Yellow Pages and competing telephone directories of merchants in the US are given away to consumers; all the revenue is provided by merchants.[20] Similarly, Microsoft and Apple do not charge applications software developers anything for the highly valuable software services (sometime called APIs) included in their software platforms. Both these firms make almost all of their money from end-users of computer systems. (In Microsoft's case, Windows is usually licensed to computer makers that in turn license it to end-users.) On the other hand, makers of video game consoles sell them to end users at or below cost and make most or all of their gross margin from license fees paid by game developers.[21]

B. Unitary Payment Systems

The general literature on multi-sided markets has immediate application to the analysis of unitary payment systems—like the BankAmericard system before franchising, the American Express

[17] In the Rochet-Tirole model, which is based on particular assumptions about the structure of demand, the price structure is completely independent of the relative costs of serving the two groups (there are no fixed costs on either side in their formulation). See Rochet and Tirole 2003b, pp. 996-997.

[18] We define gross margin here and below as revenues minus direct, side-specific variable costs, such as the manufacturing costs of video game consoles. Attributing profits to one side or another is strictly incorrect, since both sides must be on board for profits to be earned, and many costs in platform businesses, like the cost of developing the Windows operating system, are properly thought of as joint between the two sides.

[19] Bolt and Tieman 2003 offer an explanation based on the fact that some demand functions that are well-behaved in an ordinary one-sided context yield non-concave profit functions in some two-sided models, thus making corner solutions (e.g., prices of zero) optimal.

[20] In the payment card context, Vickers 2005 has argued that one might find "a distortion of competition" when one side of the market is charged "more than all the costs" of providing services to both sides. There is no theoretical basis for this test of which we are aware, and this sort of pricing is not uncommon. It occurs whenever, as in the two examples in the text, one side of a profitable two-sided business pays nothing, so the other side must be charged "more than all the costs" if the enterprise is to be profitable.

[21] See Evans et al. 2004.

system today (ignoring recent franchise-like bank deals), or the store-specific cards offered by such merchants as Neiman-Marcus. Neglecting for the moment the distinction between access and variable prices, the price structure here can be described by the ratio of fees paid by merchants, typically in the form of the merchant discount, to transaction-related fees paid by consumers, in the form of annual and other fees.[22] Using this measure, the available evidence indicates that unitary systems have generally adopted asymmetric pricing structures and earned the bulk of their revenue from merchants, rather than consumers.

Before 1950, payment cards were issued by retailers for use in their stores. Then, as now, cardholders who paid their bills within a specified amount of time (usually a bit less than a month) did not pay any fee for charging and in fact benefited from the float. Those who financed their store card charges paid interest, of course. But to our knowledge, then as now, retailers did not cancel cards held by customers who chronically did not finance. Although we would not want to push this point too far, given that store cards are one-sided and are bundled with finance services, store card transaction services have what might be thought of as a slight negative variable price and a zero access price.[23] Merchants with retail cards presumably find that this is the optimal pricing scheme.

Diners Club introduced the first two-sided payment platform—that is, a general purpose payment card that could be used by cardholders at many retailers—in 1950. After initially offering the cards for no fee to consumers, Diners Club settled on a business model in which cardholders paid an annual fee of $3 (over $18 in 2004 dollars) and a slightly negative transaction fee in the form of float, while merchants paid a variable fee of seven percent of each transaction. During the 1950s and into the early 1960s, based on available data, Diners Club earned about 70 percent of its revenues—and probably most of its gross margin—from the merchant side of the business.[24]

American Express entered the card business in 1958. It adopted a similar business model, though it eventually settled on slightly lower merchant fees and slightly higher cardholder fees than Diners Club. But available data indicate that for the last four plus decades American Express has earned upwards of 65 percent of its transaction-related revenues from merchant fees (in later years some American Express cards began to bundle borrowing with transaction services, which makes clean

[22] By "transaction-related," we mean that we are here following almost the entire literature and focusing on the payment function of credit, charge, and debit cards, putting to one side the credit function that only credit cards perform, and the associated revenue flow from consumers.

[23] Store cards may increase purchases by relaxing short-run liquidity constraints. As we note below, this argument has been advanced as a motivation for retailer acceptance of general purpose payment cards (particularly credit cards) by Chakravorti and To 2003. See also Chakravorti and Emmons 2003 and Wright 2000. However, again note that we would expect that retailers would take efforts to weed out transactors if that were the entire motivation.

[24] Evans and Schmalensee 2005, pp. 54-55. Annual fee converted to real 2004 dollars using the GDP implicit price deflator from United States Department of Commerce 2004.

comparisons more difficult).[25] Bank of America also entered in 1958, with a card that did not charge consumers a fee, although there were finance charges (and associated costs of funds and default risk) from consumers who chose to revolve. Their merchant fee was 5 percent of each transaction.[26]

Even though transactions-related costs on the acquiring side of the business seem to belower than those on the issuing side,[27] these unitary systems apparently concluded that the profit maximizing price structure for the system—the price structure that gets and keeps both sets of customers on board and permits both issuers and acquirers to be profitable—is one that obtains the bulk of the revenue from the merchant side of the business.

These unitary systems instituted these pricing schemes in their early years, when they would appear to have had little market power. We think it is fair to conclude that the"competitive"—certainly in the sense of non-collusive and non-monopolistic—pricing structure for payment cards is (or at least was for some time) one in which merchants pay a relatively high transaction price and cardholders pay zero or possibly slightly negative transaction prices plus modest fixed fees, and in which the bulk of the profits, loosely speaking, thus flow from the merchant side. Being the competitive pricing structure does not necessarily mean that it is socially optimal, however, and it is to that issue we turn next.

C. Profit versus Welfare

How do prices charged in two-sided markets compare with socially optimal prices? Most theoretical analysis of this question considers a single firm with some market power (i.e., it faces demand curves that slope down—though they may be highly price-elastic) selling to two customer groups in an otherwise perfectly competitive economy. Under these assumptions it is clear that the price level will be too high because of the exercise of market power—a conclusion that, of course, applies to most firms in real economies.

In fact, the monopoly pricing problem is even more serious than usual: even if marginal costs are constant and there are no fixed costs, the first-best social optimum requires setting prices that do not cover total cost, just as is the case with natural monopoly.[28] The intuition is straightfor-

[25] Evans and Schmalensee 2005, pp. 57-59, 150. In 2004, about 84 percent of American Express's card transaction-related revenues came from merchants, excluding finance revenues from cardholders. The proportion is about 63 percent when net finance revenues are included and using American Express's card revenues on a "managed" basis (which includes revenues for card loans that have been securitized); this same measure is 71 percent on a GAAP basis (which excludes finance charges on securitized card loans). Figures calculated using data from American Express 2005, pp. 45-49.

[26] Wolters 2000, p. 331.

[27] The portion of the merchant discount retained by acquirers on MasterCard and Visa transactions in the United States averages about 0.4 percent, which goes to cover acquirers' costs. This compares to an average interchange fee of about 1.7 percent, which, in combination with additional cardholder fees, goes to cover issuers' costs. Evans and Schmalensee 2005, pp. 11, 262.

[28] This is explicitly shown in Armstrong 2004 and Bolt and Tieman 2004b. It is implicit in Rochet and Tirole 2003b, since their analysis of socially optimal pricing imposes the binding constraint that the firm just break even.

ward. In an ordinary one-sided market under the usual assumptions, buyers' willingness to pay for incremental units of output provides a measure of the social value of that output. Thus the social optimum in such markets occurs at the output level at which price is equal to marginal cost, since at lower levels of output buyers are willing to pay more than marginal cost for incremental output, while at higher levels of output they are willing to pay less than marginal cost. In a two-sided market, however, increases in output on side A of the market provide positive benefits to buyers on side B that are not reflected in the side A demand curve. Thus if price equals marginal cost to customer group A, it is nonetheless socially beneficial to increase output to A because of the (externality) benefits that would thereby be conferred on members of group B.

What about price structure? In general single-firm models of two-sided markets in otherwise perfectly competitive economies, one can compare the conditions defining profit-maximizing pricing, welfare-maximizing pricing (which involves the seller losing money and, presumably, being subsidized by the government), and Ramsey pricing (which involves maximizing social welfare subject to the constraint that the firm does not lose money). Comparing these conditions in various models yields two general observations that appear to be robust to modeling assumptions. First, neither welfare-maximizing pricing nor Ramsey pricing is ever purely cost-based. In both cases, the optimal price structure also depends on price sensitivities and externalities on both sides of the market.[29] Second, there is no simple, general description of the relations among the profit-maximizing and Ramey price structures. They are rarely identical, but the sign and magnitude of the difference between them depends on essentially all the demand, cost, and externality parameters in the model.

III. Interchange Fees in Multi-Party Systems

Multi-party systems emerged in the mid 1960s.[30] One was the BankAmericard franchise system that we mentioned earlier. Others were cooperatives of banks that agreed to collaborate on a card brand and, in effect, pool the merchants they had signed up so that any individual with a card from a member of the cooperative could use their card at any merchant also signed up by any member of the cooperative. The predecessor of MasterCard emerged during this period as a national cooperative of banks. BankAmericard adopted a cooperative structure a few years later.

These banks initially had a purely practical problem to solve. When a cardholder serviced by bank A presented her card to a merchant serviced by bank B, the two banks had to have agreement on many issues in order to execute the transaction, even if that meant following custom or a default rule. One set of issues concerned which party bore various risks—nonpayment by the cardholder,

[29] In fact, in the Rochet-Tirole model, which assumes a particular demand structure, the Ramsey price structure, like the profit-maximizing price structure, is completely independent of group-specific marginal costs. See Rochet and Tirole 2003b, pp. 997-998.

[30] This discussion in this section is based on Evans and Schmalensee 2005, pp. 153-156.

non-delivery of goods by the merchant, bankruptcy of the merchant or, for that matter, of the merchant's bank. Another issue was how much the issuer and acquirer were compensated for executing a transaction that could not take place without participation by both of them.

As noted above, when Bank of America began franchising, its first response to this problem was a rule stipulating that whenever acquiring bank A and issuing bank B were different, A was to send the full amount of its merchant discount to B (and, of course, B was to reciprocate when one of its merchants dealt with one of A's cardholders). Presumably the idea was that if all banks adopted similar price structures, then to a first approximation typical bank A's merchant/consumer revenue mix would not depend on whether its cardholders dealt with its merchants or with those acquired by other banks. If this had worked, the system could have maintained a merchant-centered (or any other) price structure. But, as we observed above, this device failed—both because it reduced everyone's incentives to sign up merchants and because it invited deception.

In the early cooperatives, some banks entered into bilateral agreements with each other; this was possible when the systems had few members. But ultimately the cooperatives decided to develop a default set of rules, or contracts, that defined the allocation of risks and payments. In NBI—the cooperative that evolved from the BankAmericard franchise system and was the predecessor of Visa—as long as an acquirer's merchant met certain terms, such as properly authorizing transactions and checking card numbers against a list of known fraudulent accounts, it was guaranteed payment. If a transaction turned out to be fraudulent or a consumer failed to pay, the issuer was responsible. Various procedures were also set up to resolve disputes between merchants and cardholders as to the validity of particular charges. MasterCharge—MasterCard's predecessor—developed a similar contract with a similar interchange fee.

Both contracts were presumably the result of a bargain struck within the cooperative organizations between members with different stakes in acquiring and issuing (although back then there was less specialization on this dimension than there later came to be). Given the payment card industry at the time, the price structure that resulted from these interchange fees was similar to that of the unitary systems, although the price levels were lower. Since the interchange fee was a cost to the acquirer it was passed on to merchants as part of merchant discounts. The resulting merchant discounts were lower, however, than those charged by American Express at the time. The issuers then chose to issue cards with modest access fees and slightly negative transaction fees for cardholders.[31] Unlike the charge-card systems, of course, the credit cards came bundled with longer-term financing. For transactions, the resulting percentages of revenue that the multi-party systems earned from the merchant side were similar to those of American Express and Diners Club, if not higher.

[31] At least through most of the 1970s, annual fees for Visa and MasterCard cards were relatively uncommon. National Bancard Corporation v. Visa U.S.A., Inc., p. 27.

A. Alternatives to Interchange

What might the early multi-lateral card systems have done instead? Some have argued that for competition policy reasons, bilateral negotiations should have been used to set the terms of two-bank transactions rather than collective action.[32] This argument seems farfetched for the early cooperatives. They had no collective market power by any measure, and it seems more plausible that the interchange fee was, as advertised, devised to reduce the transactions costs of entering into bilateral negotiations. Moreover, as we noted above, the *NaBanco* court found that bilateral negotiations were not a practical solution in systems with many banks, both because of transactions costs and because the honor-all-cards gave issuers substantial leverage over acquirers.[33] As long as an honor-all-cards rule is in effect, so that merchants are required to accept all cards of a given brand, an acquirer is at a significant disadvantage in negotiations with other issuing banks, since its merchant is required to accept their cards, but it has no guarantee of payment by the card issuer. A guarantee of payment is possible only when the terms of payment, including the interchange fee, if any, are specified. Because of this asymmetry, there is no reason to believe that bilateral negotiation would generally lead to lower average interchange fees or merchant discounts than multilateral action at the association level.

Others have argued that the interchange fee should simply have been set at zero by competition authorities or some agency. Credit card paper would then exchange "at par," like checks in the United States.[34] One can raise several questions about this proposal. First, a zero interchange fee would result in lower prices for merchants (since acquirers would not have the interchange fee cost) and higher prices for cardholders (since issuers would not have this source of revenue). Although this might seem "fair" in a philosophical sense, there is no basis in the economics of two-sided industries for presuming that this pricing structure is more—or less—efficient than one that, like the structures adopted by the unitary systems, imposes higher prices on merchants. That is, without further information there is nothing economically special about an interchange fee of "0"; there is

[32] E.g., this was one of the remedies proposed by NaBanco. National Bancard Corporation v. Visa U.S.A., Inc. 1984, p. 1241.

[33] A recent analysis (Small and Wright 2000) concludes that, at least under certain somewhat special assumptions, even if transactions costs were low enough to make bilateral negotiations practical in large systems, banks' strategic behavior would undermine the system's viability. It is worth noting, though, that the terms of two-bank transactions in the Australian EFTPOS system (a PIN-based debit card system) are set exclusively by bilateral negotiation; see Reserve Bank of Australia 2004. There are only 11 acquirers and 150 issuers in the EFTPOS system; these numbers, while nontrivial, are orders of magnitude lower than the thousands of issuers and (at least) hundreds of acquirers in, e.g., the Visa system in the United States. See Reserve Bank of Australia 2005a, pp. 4, 40-45. Wright and Weiner 2005 report that interchange fees are also set by bilateral negotiation in Sweden. Banking concentration is quite high in Sweden, with the top four banks accounting for over 80 percent of total assets in 2003. See Swedish Bankers' Association 2004. By contrast, the top four banks in the United States accounted for only a 24 percent share in 2002. See United States Census Bureau 2004.

[34] See, e.g., Frankel 1998 and Balto 2000 and, for a rebuttal, Ahlborn et al. 2001.

no economic basis for concluding that an interchange fee of "0" is better or worse for society than any randomly chosen positive or negative percentage.[35]

Second, setting the interchange fee at zero imposes a particular price structure on the system, one in which side-specific prices are tightly linked to side-specific costs. For example, had the predecessors of MasterCard and Visa had a zero interchange fee in the early 1970s, they would have had to raise card fees by $4.88 per account (or about $18 in 2004 dollars) to compensate for the loss of interchange fee revenues.[36] Imposing less pricing flexibility on the emerging cooperative card systems in the early 1970s would have necessarily placed them at a competitive disadvantage relative to the more established unitary systems, which could choose their price structures without constraints. This distortion in competition between unitary and multi-party systems would need to be weighed against whatever benefits policymakers believe would arise from mandating exchange at par.

One also has to consider what other changes might result from a zero interchange fee. It is possible that a multi-party system is not even viable with the pricing structure that would result from a zero interchange fee.[37] In their early days, the viability of these systems was very much an open question. Even today, large numbers of issuers might move to other card systems (possibly new ones) that are organized so as to be able to replace interchange revenue without regulatory or antitrust scrutiny. (Consider, for instance, franchise arrangements with for-profit unitary systems in which merchant discounts and payments to issuers are set unilaterally by the system.) It is not clear that the pricing structure from such systems would be more favorable to merchants. Another possible effect of a zero interchange fee would be to change the other terms of the contract among member banks. For example, rules governing disputes among issuers and acquirers might be made more favorable toward issuers in order to avoid issuer defections. Or the circumstances under which payment is guaranteed to acquirers might become more limited.

B. The Baxter Analysis

William Baxter addressed the issue of system viability in his pioneering analysis of interchange fees.[38] His was also the first paper of which we are aware that showed an understanding of the two-sided nature of payment systems—or, indeed, of any market. As we will see, although his model is special and unrealistic in some respects, it provides important insights.

---

[35] Interchange payments in the Australian EFTPOS system flow from issuers to acquirers; see Reserve Bank of Australia 2004. Negative interchange fees of this sort existed in some online debit card systems in the United States until 1997 and remain the norm in U.S. ATM systems; see Hayashi et al. 2003.

[36] Calculated from data provided by Visa U.S.A. Card fee converted to real 2004 dollars using the GDP implicit price deflator from United States Department of Commerce 2004.

[37] We illustrate this point in a simple model immediately below.

[38] Baxter 1983.

Baxter assumed perfect competition among issuers, among acquirers, and everywhere else in the economy, and, like almost all the subsequent literature, he assumed away all fixed costs and access prices. Under these conditions, collective determination of the interchange fee cannot be an exercise of market power, since there is no market power anywhere in the economy. It is simply a payment from one set of perfectly competitive firms, which will have to raise their (variable) prices to cover it, to another set of perfectly competitive firms, which will lower their (variable) prices so as to compete it away completely. The interchange fee thus can only affect the price structure, not the price level. Baxter assumed that consumers and merchants would use a particular payment card if and only if the per-transaction price charged to them was less than the per-transaction benefit from using the card rather than cash or check.

To see the role of the interchange fee in ensuring system viability in this setup, suppose for simplicity that all consumers have the same per-transaction benefits and that so do all merchants. Let the per-transaction prices charged by the card system to consumers and merchants be P^c and P^m, respectively; let the corresponding per-transaction benefits relative to cash or check be B^c and B^m; and suppose the constant per-transaction marginal cost of serving a consumer is C^c and of serving a merchant is C^m. Under these assumptions, it is efficient to use the card for all transactions if and only if

$$(B^c + B^m) \geq (C^c + C^m). \tag{1}$$

With a zero interchange fee and perfect competition, consumers will agree to use the card and issuers will break even if

$$B^c \geq P^c = C^c, \tag{2}$$

And merchants will agree to accept the card and acquirers will break even if

$$B^m \geq P^m = C^m. \tag{3}$$

For a zero-interchange system to be viable, both (2) and (3) must be satisfied. It is easy to find numerical examples in which (1) is satisfied, but either (2) or (3) is not. Suppose, for instance:[39]

$$B^c = 1,\ B^m = 8,\ C^c = 3, \text{ and } C^m = 2. \tag{4}$$

Here total per-transaction benefits from using the card are almost double the corresponding cost, but there is no price to consumers that satisfies (2). In this example retailers receive the most benefits if the card is used, and it would be logical for them to cover most of the system's costs. But there is no way to accomplish this with a zero interchange fee. An interchange fee of 2 solves the

[39] This example is from Rochet 2003, where this model is instructively explored in more detail.

problem. This raises the acquirer's cost and (because of perfect competition among acquirers) the merchant discount, P^m, to 4, and it lowers the issuer's cost and (because of perfect competition among issuers) the consumers' fee to 1. Merchants contribute 80 percent of the system's revenue but are better off by 4 per transaction, while consumers are just indifferent to the card's existence even though the positive interchange fee has reduced their fees substantially.

By assumption, the costs incurred by both sides are necessary to execute a transaction. If this assumption is correct, it makes no sense to think of either side as providing particular services to the other; both must incur all the costs stated for either to benefit. Moreover, there is no way for regulators to look only at cost conditions and conclude that an interchange fee of 2 is appropriate. Even in this simplest possible example, demand conditions must be considered. It could happen, of course, that some cost-based formula produced an interchange fee of 2. But this could only happen by chance.

In Baxter's model, the interchange fee is not set to maximize profits, since there are no profits earned anywhere in the payment system. He argued that its level was determined uniquely by the need to balance the supply and demand of card transactions. This seems somewhat artificial, since as long as some merchants have agreed to accept a card, the volume of transactions is determined unilaterally by card-carrying consumers' deciding whether or not to use the card for particular transactions. Nonetheless, the proposition that at some interchange fee levels the system would not be viable is likely to hold in more general models. The important—and robust—insight from Baxter's analysis, although it is not framed precisely this way, is that the interchange fee helps internalize an externality between the two customer groups and in so doing has the potential of making both customer groups better off.

C. Imperfect Competition in Issuing and Acquiring

The assumption of perfect competition between homogeneous issuers and acquirers is not realistic and leads to most interesting questions having indeterminate answers—since there are no profits to be had, there is no motivation for doing anything. The next logical step was taken by Richard Schmalensee.[40] He made the standard assumption that, except for the payment system under study, the economy was perfectly competitive.[41] This implies that the demand system facing the system, which Schmalensee took as given and did not derive from first principles, could be used for standard welfare analysis. Schmalensee allowed for imperfect competition among issuers and/or among acquirers and made a particular assumption about the functional form of the demand system and thus about the structure of indirect network effects. As in the Baxter analysis, the system itself was

[40] Schmalensee 2002.

[41] Much of the subsequent literature has persuasively called that standard assumption into question in this context, as we discuss in the next section. Nonetheless, it is instructive to see its implications before considering how it might be altered in the direction of greater realism.

assumed to operate like the Visa and MasterCard systems operate in fact, on a break-even basis, and fixed costs and access prices were again assumed away. Thus, as in the numerical example above, the interchange fee simply shifts costs between issuers and acquirers, raising costs on one side of the market by exactly as much as it lowers costs on the other side.

In this model the level of the interchange fee can affect the profits earned by issuing and acquiring banks whenever competition among them is imperfect, making it possible to compare profit-maximizing and welfare-maximizing fees. Under some special assumptions the comparison is simple: when there is a single issuer and a single acquirer and demand curves are linear, for instance, Schmalensee shows that the profit-maximizing interchange fee also maximizes system output and economic welfare. In this case regulation could only reduce overall performance. In general, however, these three fees may be different even under the particular demand structure Schmalensee assumed, and even when the further assumption of linear demand is imposed. Thus, even under strong assumptions about demands and costs, the relations among these quantities is complex and depends on demand parameters, cost conditions, and, an element not present in analysis of unitary two-sided platforms, the nature of competition among issuers and among acquirers.

In the Schmalensee model the interchange fee is not an ordinary market price: it is a balancing device for shifting costs between issuers and acquirers and thus shifting charges between consumers and merchants.[42] Fixing the interchange fee is quite unlike fixing a price in a typical one-sided market. The first-order effect of ordinary price-fixing is to harm consumers by restricting output. The first-order effect of collective determination of interchange fees in this model is generally to enhance the value of the system by balancing participation of the two customer groups, thus internalizing indirect externalities. This is illustrated most clearly by the existence of a special case, noted above, in which collective determination of the interchange fee in order to maximize profit also maximizes output and economic welfare. And even in Schmalensee's simple model, the socially optimal interchange fee depends on costs, demand conditions, competition among issuers and among acquirers, and externalities between merchants and consumers. Thus as a practical matter, there is no rule for regulatory determination of the interchange fee that could be relied on to improve overall system performance and thus enhance economic efficiency.

IV. Second-Best Interchange Fee Analysis

Except for Schmalensee, most writers on interchange assume some market distortion in addition to imperfect competition among issuers and/or acquirers. In these models social welfare analysis cannot generally be based on the demand system facing issuers and acquirers, and that demand system is derived from more fundamental assumptions rather than assumed as in the Schmalensee paper. These authors then generally examine how the additional distortion they consider affects

[42] In addition, the interchange fee can be used to shift profits between issuers and acquirers, though when it is used for that purpose total system profit is necessarily reduced.

the relation between the profit-maximizing and welfare-maximizing interchange fees. Almost all assume away fixed costs and access pricing and assume that all revenues flow from variable prices. When used to consider regulatory policy, these models become exercises in the economics of the second best.[43] That is, they consider policy in the presence of multiple, interacting departures from the competitive ideal. In general, such exercises rarely yield tractable rules that can be used to design practical policies. The interchange literature, which points to many additional distortions but typically considers them one at a time, provides no exceptions.[44]

A. Imperfect Competition Among Merchants

In a very influential paper that was the first analysis of interchange to derive system demand functions from first principles, Jean-Charles Rochet and Jean Tirole allowed for imperfect competition among merchants.[45] They assumed perfect competition among acquirers and, for simplicity, identical merchants. They also assumed that some fraction of consumers was more likely to patronize merchants who accepted cards, so that merchants had a strategic incentive to accept cards in order to avoid losing the economic profits they would earn from selling to those consumers. Thus, in contrast to Baxter's model, because of this strategic, rent-seeking incentive, merchants would find it optimal to accept cards even if transactions using them were somewhat more expensive than transactions using cash.

Since all profits in this model system are earned by issuers, the profit-maximizing interchange fee is the highest fee consistent with all merchants accepting the card. This may be equal to the socially optimal fee, but it may be higher because competition forces merchants to internalize part of the benefits of cardholders. In the former case card usage is welfare-maximizing, but in the latter case it is excessive, since an interchange fee that is too high drives consumer variable fees too low and thus stimulates excessive card usage by consumers.[46] No cost-based or other simple rule for regulating the interchange fee reliably solves this problem, however, even in this bare-bones model.

The result that a profit-maximizing system never sets an interchange fee below the welfare-maximizing level in the Rochet-Tirole model, which seems to underpin some current regulatory initiatives, depends critically on the very strong assumption that merchants are identical. Julian Wright shows that if this assumption is relaxed and even if consumers know which merchants accept cards (so merchants' rent-seeking incentive to take cards is maximized), the profit-maxi-

[43] Lipsey and Lancaster 1956.

[44] For surveys of some of the literature discussed here, see Chakravorti 2003, Hunt 2003, Rochet 2003, Rochet and Tirole 2003a, and Schmalensee 2003.

[45] In an influential review of the literature undertaken for the Reserve Bank of Australia, Katz 2001 stressed the effects of such competition. See also Wright 2003c and Hayashi 2004.

[46] It is worth noting that De Grauwe and Rinaldi 2002 have developed a model involving rent-seeking by merchants in which the level of card usage is below the welfare-maximizing level. They treat merchants' decisions to accept or not accept cards as exogenous, however.

mizing interchange fee may be above or below the welfare-maximizing level, and there may thus be too many card transactions or too few.[47] The relation between profit-maximizing and welfare-maximizing fee levels in this more general model is complex and depends on details of demand and competitive conditions as well as costs.

B. Reducing the Transactions Costs of Borrowing

Credit cards bundle the provision of credit with transaction processing. In so doing they reduce the transactions costs of borrowing on the part of cardholders, so that cardholders can buy on credit more easily. Through the development of sophisticated risk scoring methods, credit cards may also have relaxed the overall liquidity constraints that many consumers face. In addition, charge cards provide short-term liquidity, albeit only for a couple of weeks on average, and debit and charge cards reduce the transaction costs of obtaining funds from alternative sources. If a merchant's decision to accept cards induces consumers sometimes to spend more than they otherwise would because of these reductions in consumers' transactions costs, merchants have another incentive to accept payment cards, over and above their own transaction cost savings and the rent-seeking incentive discussed above.[48] There is also at least anecdotal evidence that these are benefits that merchants receive at lower cost from general purpose payment cards than they could have provided themselves through store card programs. Many smaller retailers dropped store cards as credit cards became more widely held, and few retailers steer customers to their own store cards anymore.

It is hard to evaluate the welfare implications of providing increased liquidity. Relaxing liquidity constraints and reducing transaction costs clearly benefit consumers. The more merchants that accept cards, the larger these benefits. A rigorous analysis would need also to consider the sources of both liquidity constraints and transaction costs, however, and we are not aware of any empirical or theoretical analysis that even attempts to do this.

C. Competition Among Payment Systems

In the United States, payment card systems compete against cash and checks as well as each other. The production of cash is a government activity, subsidized through the Federal budget. And cash users do not fully internalize some of the social costs of using it, such as crime against merchants. The check system in the US is run by the Federal Reserve, which essentially forced banks early in the last century to exchange checks at par—that is, to have a zero interchange fee in the checking system.[49] The price structure in this competing system is thus not fully market-determined. Similarly, outside the United States, governments have commonly influenced the evolution of Giro

[47] Wright 2004a.
[48] Chakravorti and To 2003. See also Chakravorti and Emmons 2003 and Wright 2000.
[49] Chang and Evans 2000; for an alternative view, see Frankel 1998.

(a payment system in which a bank or a post office transfers money from one account to another when it receives authorization to do so) and related transactional systems.

There is no rigorous analysis of which we are aware of the effects of government-determined pricing in these competing systems on pricing in payment card systems. It is worth noting, for instance, that a large fraction of consumers in the United States and Western Europe pay no variable fees for writing checks, primarily as a result of the decision by banks to bundle this service (and ATM/debit cards) with the general banking relationship. This might explain why credit and charge card systems have generally not imposed transaction-specific variable fees on consumers. Similarly, there has been no analysis of which we are aware that considers the impact of this sort of competition on the relationship between profit-maximizing and welfare-optimizing interchange fee levels.

The existing literature does contain a number of (generally complex) analyses of competition between payment systems or between two-sided platforms in general.[50] It is fair to say that this work is at an early stage. It seems clear that the nature of consumer and merchant behavior shapes the competitive price structure, and that there is no general tendency for competition between platforms to make the price structure closer or farther from the social optimum, though competition will generally tend to lower the price level. Suppose, for instance, that consumers all have multiple cards (this is termed "multi-homing" in the literature) and are indifferent among them, and that merchants can effectively persuade consumers which card to use. In this case competing systems will have an extra incentive to compete for merchants' favor, and this will cause a tilt in the price structure against consumers that may or may not improve performance. Similarly, if consumers tend to use only one card (to "single-home") while merchants find it easy to accept all cards, competing systems will have an extra incentive to attract consumers, and the price structure will be tilted in their favor accordingly. These aspects of consumer and retailer behavior have received little empirical study, however.[51]

D. Barriers to Surcharging

In the United States and, at least until recently, elsewhere, payment card systems seem to have generally required merchants that accept their cards to agree not to impose a surcharge on consumers who use those cards. At first blush, this no-surcharge rule (NSR) would appear to be an artificial distortion likely to reduce performance. Dennis Carlton and Alan Frankel were the first to observe that, in a fully competitive system (*a la* Baxter), if there is no NSR, and if it costs merchants nothing to charge different prices to consumers depending on what payment system they

[50] See Armstrong 2004 and 2005, Armstrong and Wright 2004, Chakravorti and Rosson 2004, Gabszewicz and Wauthy 2004, Guthrie and Wright 2003, Manenti and Somma 2003, Rochet and Tirole 2002 and 2003b,

[51] Rysman 2004 finds that most U.S. consumers carry multiple cards but tend to concentrate their purchases on only one, suggesting a weak form of single-homing. It is unclear what combination of preferences, costs, and merchant behavior this finding reflects, however, or how much a typical consumer prefers her primary card over others she carries.

use, the interchange fee will be irrelevant and card usage will be efficient.[52] If acquirers are perfectly competitive, an increase in the interchange fee is passed along dollar for dollar to the merchant discount, and perfect competition among merchants means that the merchant discount is passed along dollar for dollar to card-using consumers (and not at all to those who pay with cash). On the other side of the market, competition among issuers means that the increase in interchange is passed to card-users, dollar for dollar, in reduced fees. Thus card-using consumers pay the full cost of the system regardless of the interchange fee, and, if they also bear the full costs of all other payment systems, they will use payment cards if and only if they are socially less costly. Accordingly, Carlton and Frankel advocate abolishing NSRs.

It turns out that the Carlton-Frankel assumptions are stronger than necessary for the interchange fee to be irrelevant in the absence of an NSR. What is required is only costless surcharging—i.e., the ability of merchants at no cost to charge different prices depending on the means of payment used.[53] But without perfect competition everywhere, abolishing an NSR does not generally lead to an efficient outcome. In particular, imperfect competition among issuers then tends to lead to under-provision of card services, and merchants could use surcharges as a mechanism for price discrimination. Economic welfare may be lower than at the profit-maximizing equilibrium with an NSR—even if card usage is excessive in the latter case. In the murky realm of the second-best, this sort of ambiguity is not uncommon.

Moreover, it is clear that the assumption of costless surcharging is unrealistically strong; most merchants do not discriminate among people using different means of payment even when they are not prevented from doing so. For instance, it has generally been permissible for U.S. merchants to give a discount for cash purchases and, though this was done for a time at gasoline stations, it is now extremely rare.[54] In the Netherlands about 10 percent of merchants imposed surcharges when they were allowed.[55] In the United Kingdom, surcharges are permitted for credit and charge cards but surcharging is uncommon.[56] And when NSRs were abolished in Sweden, only about 5 percent of merchants imposed surcharges.[57] Although we are not aware of any concrete data yet from the elimination of NSRs in Australia,[58] our understanding from colleagues there is that the prevalence of surcharging, at least to date, is likely in line with these other experiences. The one instance we are aware of with a somewhat higher incidence of surcharging, although still far from pervasive,

---

[52] Carlton and Frankel 1995a and 1995b; see also Evans and Schmalensee 1995.
[53] See Rochet and Tirole 2002, Wright 2003a, and, especially, Gans and King 2003b. Schwartz and Vincent 2004 provide a related analysis for a unitary system.
[54] Chakravorti 2003, p. 55.
[55] ITM Research 2000.
[56] Office of Fair Trading 2003b, pp. 114-115.
[57] IMA Market Development AB 2000, p. 18; but see Katz 2001, p. 44.
[58] For example, the Reserve Bank of Australia notes that "some merchants" have started surcharging but that "there are no comprehensive data." Reserve Bank of Australia 2004, p. 11.

was in Denmark, where earlier this year, 19 percent of merchants (primarily grocery retailers) passed a new 0.55 Kroner debit card surcharge through to consumers.[59]

We suspect merchants are reluctant to impose surcharges for two reasons. The first is that there are transaction costs of imposing different prices based on payment methods. The second is that consumers may prefer to patronize stores that do not surcharge. No analysis of which we are aware has considered or attempted to measure these costs and preferences. There is nothing unusual here, of course: there are many things that merchants could surcharge for—because they entail specific costs that are caused by particular customers—and do not.[60] Parking in shopping malls is an obvious example.

E. Implications

The quantity of recent theoretical literature discussed in this section makes it clear that economists find interchange fees fascinating. Almost all of these papers find that profit-maximizing interchange fees are unlikely to be socially optimal, but none yields workable rules for welfare-improving regulatory intervention. (In particular, none points to the optimality of any rule that is purely cost-based.) Moreover, these models are highly stylized. A variety of market imperfections are considered one at a time; not only are some visible imperfections not considered at all (e.g., government-determined zero interchange in U.S. checking), no analyst has even attempted to consider them all together. Most papers assume consumers are faced with variable (i.e., transaction-specific) charges, even though most systems do not impose such charges and rely instead on annual fees and other access charges. Finally, there has been essentially no empirical work devoted to testing any of these models or to measuring the importance of any of the effects they predict. This literature, in short, is not very useful for either rationalizing or designing a system of interchange fee regulation. Of course, for exactly the same reasons, it is not capable of proving that the interchange fees determined by the card associations are exactly or approximately socially optimal.

It is useful to compare the results for interchange fees for payment cards with some other business practices that economists have analyzed. In many ways the theoretical results surveyed here are similar to those for advertising, research and development, product design, product variety, location decisions, firm entry in the presence of fixed costs, bundling, and price discrimination.[61] In all these cases economic models show that the profit-maximizing result under imperfect competition may deviate from the social welfare-maximizing result. However, in most cases the bias can go in either direction, and in all cases determining the socially optimal result depends on complex factors that cannot be measured in practice. Based on our review of the theoretical literature

[59] The Danish competition authority had announced that it was going to investigate whether there had been any collective action by grocery stores. Konkurrencestyrelsen 2005.

[60] See, for example, Evans and Schmalensee 2005, p. 131.

[61] See, generally, Tirole 1988.

on interchange fees to date there does not seem to be any basis for concluding that the potential distortions caused by collective determination of interchange fees are any more—or any less—significant than the potential distortions caused by these other deviations from the ideal model of perfect competition.

V. Government Determination of Interchange Fees

Traditionally, government control of prices and conditions of service—either via government ownership or economic regulation—has been most prevalent in network public utility sectors, such as water, electricity, telephone, and gas. Firms in these sectors were traditionally local monopolists. The stated purpose of regulation or government ownership in these sectors was to protect consumers from prices that would otherwise reflect the exercise of monopoly power and thus, as every student of basic economics should be able to explain, would be too high. As all of those students should also know, efficient prices in these sectors are based on marginal costs, with socially optimal markups above marginal cost depending on demand conditions. However, the global movement toward privatization, deregulation, and incentive regulation reflected an emerging consensus that even in these near-textbook cases, economic welfare was in practice not reliably improved by government ownership or price regulation.[62] Politics inevitably intrudes into government price-setting, for instance, and limitations on profit rates tend to lead to waste and inefficiency.

Experience with government control of prices has taught analysts that a persuasive economic case for price regulation requires a positive answer to two questions:

1. Is the performance of the market or markets being considered substantially suboptimal?
2. Is there a practical regulatory policy that is reasonably certain to improve market performance substantially?

Since regulation is a blunt policy instrument in practice, unless there is a substantial market failure there is scant chance that regulation will reliably improve matters. And unless there is a known, practical regulatory rule that, if followed, is reasonably certain to improve performance, it is likely that regulation will be on balance harmful. At least in theory, there is no economic point to interfering with even imperfect markets unless those imperfections are serious and capable of correction by known methods.

[62] See, for instance, Armstrong et al. 1994, Megginson and Netter 2001, and Littlechild 2003. In the United States, price cap regulation has been used as an alternative to rate of return regulation by the FCC and some states in regulating telephone companies starting in the 1980s. Viscusi et al. 2000, Chapter 12. This approach has been very widely used outside the United States. Winston and Morrison 1986 document some of the adverse effects of airline regulation in the US.

A. The First Question: Is There a Significant Market Failure?

As we have discussed, the growing body of theoretical writing on interchange fees establishes that privately optimal fees are unlikely to be socially optimal, but it does not indicate whether they will be systematically too high or too low. This literature does not and, in the absence of evidence, cannot indicate whether non-optimal interchange fees have a significant or trivial effect on overall performance. The related literature on pricing structures for two-sided markets reaches a similar result.[63] There is no reason to presume that even competing two-sided platforms will settle on a price structure that is socially optimal. And, for this reason, there is no reason to presume that unitary systems have pricing structures near the social optimum.

Empirical evidence, not theory, therefore must play the leading role in assessing whether interchange fees—and the resulting prices to cardholders and merchants—lead to significant underprovision or overprovision of payment card services relative to the social optimum. In a recent discussion of the rationale for regulation of interchange in Australia, I.J. Macfarlane, Governor of the Reserve Bank of Australia, similarly stresses the results of a factual inquiry rather than the deductions of economic theorists:[64]

> …we saw that credit cards were growing faster than the other means of payment. This was initially somewhat surprising as credit card transactions are more expensive than most other means of payment—that is, they involve a larger payment from the users of the payments system to the providers of the payments system. … Why was this possible?

Governor Macfarlane poses *almost* the right question: Are payment system usage patterns significantly inconsistent with system costs—and, it is essential to add, benefits? There is nothing unusual about a high-cost product driving out cheaper competition if the high-cost product is much better. U.S. drivers generally prefer automatic to manual transmissions in their automobiles, for instance, even though automatic transmissions cost more and are more expensive to maintain. Drivers seem to believe the difference in benefits outweighs the difference in cost.[65]

In our context, U.S. banks typically charge consumers the same variable price for handling a check as for handling a signature-based debit card payment: zero. Debit cards are nonetheless rapidly replacing checks because consumers find them more convenient.

---

[63] It is worth noting one difference explicitly. Both profit- and welfare-maximizing interchange fees depend in general on the nature of competition among issuers and among acquirers, and this additional complexity is not present in the analysis of ordinary, unitary two-sided platform businesses.

[64] Macfarlane 2005.

[65] Similarly, as Joanna Stavins pointed out in a paper on electronic check truncation and presentment, bicycles may be cheaper than cars but that does not mean that society would necessarily be better off if bicycles replaced cars. Stavins 1997, p. 28.

An approach that is based on careful measurement of costs and benefits has the potential to distinguish what is important from what is only a theoretical possibility. Dealing with quantitative evidence rather than qualitative possibilities can also inform regulatory policies if regulation is deemed appropriate. Since the objective of interchange-fee regulation, where it is warranted, should be to correct the effects of a distortion of price signals, it is important to get a quantitative sense of the importance of those effects. Let us, accordingly, turn to the available evidence.

Some studies have argued that payment cards are used too much, based on the observation that they are more expensive on average for merchants than cash and checks but are nevertheless increasing in use at the expense of cash and checks.[66] While this is intuitively appealing, it ignores both cardholders and the role of benefits in determining the social optimum.

To see this, consider a simple economy with only cash and cards and with a fixed set of transactions to be executed. For each transaction, one can in principle compute the social marginal cost of executing the transaction using cash—the sum of the marginal costs to the merchant and consumer involved, as well as the net costs to all other involved parties, including governments and commercial banks. One can similarly compute the social marginal cost of executing each transaction using cards, along with the social marginal benefit of using cards as opposed to cash—the sum of the marginal benefits to consumers, merchants, governments, and commercial banks.[67] The marginal benefit in any of these cases may be negative, of course; consumers may find cards less convenient than cash for very small transactions, for instance. In this simple case the efficient outcome is clearly for cash to be used for transactions for which the social marginal cost of using cash is less than the net social marginal cost of using cards—the social marginal cost of using cards minus the social marginal benefit of using cards rather than cash—and for cards to be used for the others. Transactions-specific marginal costs and benefits are in principle necessary to assess the importance (i.e., the net social cost) of any deviation from this ideal. Considering only costs and only one of the parties involved cannot be very informative.

To our knowledge, there are no empirical studies in the literature that consider the marginal social costs and benefits for merchants and cardholders, and thus there are no comparisons of actual versus optimal use of alternative payment systems that have even approximate economic validity. There is some evidence on marginal costs for merchants, some highly incomplete evidence on marginal benefits for merchants, and essentially nothing on marginal costs or benefits for cardholders or other parties.[68]

Garcia Swartz, Hahn, and Layne-Farrar provide a useful overview of much of the available evidence, examine some of the issues one would need to consider in evaluating whether an economy

[66] See, for example, Balto 2000 and Reserve Bank of Australia 2002.

[67] For consumers, benefits would be measured by changes in consumers' surplus; for the other entities the measure would generally be changes in rents.

[68] Compare, for instance, Humphrey et al. 2003, which argues that electronic payments systems are inherently socially more efficient, and Food Marketing Institute 1998, which argues that payment cards impose excessive costs on grocers.

has roughly the socially optimal use of payment cards, and present some rough calculations based on available data and some plausible assumptions about consumer benefits for a few transactions sizes and types.[69] They take the merchant-based cost surveys as a point of departure but then proceed to make three types of adjustments. First, merchant-based surveys of costs usually compare payment instruments at different transaction sizes, typically the average transaction size for each payment instrument. The authors depart from this norm and compare payment instruments at a set of fixed transaction sizes, small and large, to assess the sensitivity of their results. Second, by combining plausible assumptions with available data they attempt to incorporate all other parties to the transaction—consumers, the government, and commercial banks—into the calculation of cost and benefits. Finally, they consider the benefits received by the cardholder and merchants; as a result they compare the "net" cost of payment methods. Garcia Swartz, Hahn, and Layne-Farrar present rough estimates of these magnitudes and find that the optimal payment method (considering quantifiable costs and benefits) varies by transaction size, store type, and other circumstances. For many transactions it appears that credit and debit cards do not have higher "net costs" than cash or checks. While this work is instructive, however, it falls well short of a rigorous overall assessment of payment system performance.

B. The Second Question: Will Interchange Regulation Help?

Let us assume that one has shown that it would be socially optimal to reduce the use of payment cards in favor of other payment systems. Two further questions would then need to be considered. (1) Is the interchange fee the appropriate method for trying to achieve this improvement? (A related question is whether the source of the distortion is the collective setting of the interchange fee.) (2) If the answer to that question is affirmative then do we have a method for regulating the interchange fee that is likely to increase social welfare? Since we know that interchange fees can be too high or too low, and either case can lead to a distortion, the issue is whether regulators can estimate the optimal interchange fee precisely enough to have confidence that they will improve social welfare.

The first question is more problematic than it might seem at first. As we noted earlier, card issuers charge merchants mainly variable fees and charge cardholders both access (i.e., fixed) and variable (i.e., transaction-specific) fees. The claim has been that cards are over-used because issuers—benefiting from and competing away to some extent interchange fee revenue—impose too low (possibly negative) variable fees on cardholders. Cardholders, who do not bear the costs imposed on the merchant by their choice of payment method, it is thus argued, use their cards too much.

A regulation-mandated reduction in the interchange fee tends to reduce the variable fees faced by merchants; if competition in acquiring is sufficiently intense, this reduction is one-for-one. However, reducing the interchange fee does not necessarily raise the variable fee paid by cardholders

[69] Garcia Swartz et al. 2004.

one-for-one—or, indeed, at all. Issuers can respond to the loss of interchange fee revenue by varying either fixed or variable fees, or both. The extent to which they vary each will depend on a variety of factors including the elasticity of demand for access to cards and the elasticity of demand for transactions. In addition, there are likely to be one-time costs of various sorts (including costs of changing accounting and billing systems) caused by moving from the traditional regime of zero variable charges to a regime with positive variable charges. The one-time costs of making reward programs less generous are likely to be less significant, if they exist at all.

Suppose, for instance, that issuers increased annual fees but did not reduce variable fees at all. In that extreme case the regulation of the interchange fee would not alter consumer incentives to use cards, although it might reduce the average number of cards that people carry. (Since it is hard to function in modern economies without at least one payment card, we doubt that the fraction of households with zero cards would rise noticeably.) Preliminary data from Australia suggests that even though interchange fees were reduced by nearly half in late 2003, the marginal price to cardholders of using credit cards has not changed much. While some reward programs were made somewhat less generous, these cover only a fraction of consumers and card transactions,[70] and we have not seen widespread evidence of surcharging by merchants or the imposition of fees by issuers that increase with card usage.[71] So while available data suggest that interchange fee reductions were passed through more or less completely to reductions in merchant discounts, it does not seem that the stated objective of the RBA to make consumers face the "right" variable prices for different payment methods was realized to any appreciable extent.[72] Consistent with this, the data do not reveal much, if any, impact of the reforms on the use of credit cards.[73]

---

[70] Industry estimates suggest that reward cards constitute about one third to one half of all credit cards in the United States; we lack comparable data for Australia. Bayot 2003.

[71] Making some reward programs less generous moved some negative variable prices toward zero. Many of these changes were in the form of caps on the total number of rewards points that can be earned each year. These limits have been set at relatively high levels, over AUS$40,000 (approximately US$31,000 at the time of this writing) a year, and likely have not been a binding constraint for many cardholders. See Reserve Bank of Australia 2004, p. 11. There were also some reductions in the marginal reward per dollar spent in some cases. There is no reason to think that this would have a big effect on card volume. In fact, if one is accumulating points in a rewards program to take a trip, and the rewards program is made less generous so that more purchases are required to earn the trip, one might use the card more rather than less intensively. There would be costs of various sorts associated with imposing positive transaction fees on non-reward cards, so it is not surprising that this does not seem to have been done.

[72] The RBA stated that its reforms were in line with the principle that "consumers should face prices that take into account the relative costs of producing goods and services, as well as demand conditions." Reserve Bank of Australia 2002, p. 34.

[73] Credit card dollar volume grew significantly in Australia in the late 1990s. The rate of growth started declining following 1999. The rate of decline in the growth rate leveled off in 2003 and 2004, in contrast to an acceleration in the rate of decline that would be expected were the reforms to have had a substantial impact. Nonetheless, the real dollar volume of credit card purchases rose 20.0 percent between 2002 and 2004, about four times as rapidly as real household consumption. Interestingly, the number of credit card accounts grew by 11.1 percent over this period, compared to a growth of only 2.6 percent in population. See Reserve Bank of Australia 2005b.

All this is not to say that dramatic reductions in the interchange fee had no economic effects. Because retailing in Australia is relatively concentrated and thus competition is likely to be imperfect,[74] it is reasonable to expect that only a fraction of the fall in merchant discounts was passed on to consumers in the form of lower prices, and the remainder went to increasing retailers' profits. When consumers use cash or credit cards without rewards they are better off as a result; when they use cards with reward schemes they are likely worse off. We have seen no evidence that these effects can be associated with distinct groups of consumers – different income quintiles, for instance.

On the other side of the market, the best evidence we have been able to obtain indicates that issuers have recovered between a third and a half of the fall in interchange revenue through increased fees to consumers.[75] Thus issuers' profits have been reduced, and consumers with credit cards of all sorts have been directly harmed, particularly if they use reward cards. In order to mitigate reductions in MasterCard and Visa interchange revenues, three of the top four Australian banks have signed agreements to issue American Express or Diners Club cards, which can provide greater, unregulated transaction-related revenues to issuing banks.[76] In part as a consequence, the shares of these unitary systems have risen, though from initially low levels.[77]

All in all, it seems that in the short time since interchange fee reductions were imposed retailers have been made better off, issuers have been made worse off, and some consumers have been made better off (particularly those who tend to use cash a lot) and others have been make worse off (particularly those who use credit cards with reward schemes). In the long run some of these effects are likely to be undone as consumers and issuing banks move toward American Express and

---

[74] Government statistics on firm concentration are not available. We have identified what data we could find from a range of sources. Many merchant categories appear to have significant levels of concentration. For example, within their respective categories, the top department store had a 71 percent share (and the top two had an 83 percent share) in 2003; the top two supermarket and grocery stores (excluding convenience, specialty and miscellaneous food stores) had a 75 percent share in 2003; the top two mobile telephone operators had a 78 percent share in 2005; the top land-line telephone operator had a 75 percent share in 2005; and the top two airlines had an 83 percent share in 2005. In contrast, the four-firm concentration index for these categories is significantly lower in the United States based on 2002 U.S. Census data: 65 percent for department stores, 33 percent for supermarkets and grocery stores, 61 percent for mobile telephone operators, and 60 percent for land-line telephone operators. The U.S. Bureau of Transportation Statistics reports the top two firms accounting for 32 percent of enplaned passengers among major U.S. airlines in 2000. The one merchant category that we found with lower concentration was for warehouse clubs and superstores, for which the four-firm concentration index in Australia was 41 percent, compared to 92 percent in the United States. See Euromonitor 2005a, Euromonitor 2005b, Euromonitor 2005c, United States Census Bureau 2004, Bureau of Transportation Statistics 2000, Maxwell 2005, and McFarland 2005.

[75] See Chang, Evans and Garcia Swartz 2005.

[76] See, e.g., Cornell 2004 and 2005.

[77] Reserve Bank of Australia 2004, p. 13.

Diners Club, as these schemes have substantially higher average merchant discounts than the bank card systems.[78]

The apparent ineffectiveness of the fairly dramatic Australian regulatory intervention in terms of its stated goals may result from the particular competitive environment in that country and the types of cards used by consumers. However, it would appear to be generally the case that the interchange fee is a highly imprecise instrument for affecting the volume of transactions on cards and thus for correcting any perceived market distortion. That is because there is only a loose connection between interchange fees and transactions prices to cardholders.

Let us suppose for the sake of argument that regulators went further than any have gone in fact and banned the use of consumer access fees. Interchange fee regulations would then necessarily affect the transaction price paid by consumers, as they are assumed to do in the theoretical literature. (This ban would, of course, likely impose a new set of costs and distortions on the system.) There would still be the question whether it is possible to estimate the optimal interchange fee with sufficient precision that policymakers could expect to increase rather than decrease social welfare. A robust conclusion from the theoretical literature is that an estimate of the optimal interchange fee would depend on a host of factors: estimates of the price responsiveness of cardholders and merchants, indirect network effects between cardholders and merchants, competition in issuing and acquiring and among merchants, price distortions in competing payment systems, transactions costs and liquidity constraints, and marginal costs of serving cardholders and merchants. It would also depend on how competing systems—some of which may be unitary—would respond to changes in prices to cardholders and merchants. Because of the difficulty of the task, there are no serious attempts of which we are aware to estimate the socially optimal interchange fee for any real payment system. Given currently available data and estimation methods, we believe that any such attempt could at best yield highly imprecise estimates.

This task would require far more empirical information than classic public utility regulation. Most public utilities have historically been monopolies, so the strategic interaction with competitors could be ignored, unregulated prices can be safely presumed to be too high, and reducing

[78] Some have argued that reducing the interchange fees of the multi-party systems will put pressure on unitary systems to follow suit, but this is not persuasive. To the extent that interchange regulation disadvantages the multi-party systems, it reduces competitive pressure on the unitary systems. (We are indebted to Michael Katz for this point.) The Reserve Bank of Australia suggests that there has been some downward pressure on merchant discounts for American Express and Diners Club in Australia, with a decline of 13 basis points over the 12 months ending June 2004 (although it does not discuss whether other factors might have been responsible for this), but expresses concern that the gap between the merchant discount for the unitary versus multi-party systems had widened by 30 basis points. See Reserve Bank of Australia 2004, pp. 12-13. Given that the merchant discounts for the unitary systems were, according to one account, about 100 basis points higher than for the multi-party systems (Reserve Bank of Australia 2001, Figure 2.2), even if there is a modest decrease in the merchant discounts for multi-party systems attributable to the reforms, the post-reform unitary merchant discount would still exceed the pre-reform multi-party merchant discount. And the merchant discount paid on volume shifted from the multi-party to the unitary systems would increase.

prices until the utility just breaks even will generally increase economic welfare—at least as long as impacts on the utility's incentives for efficiency are ignored. Although sometimes present, network effects rarely played an important role in the analysis, and the calculation of optimal (Ramsey) prices required only estimates of marginal costs and demand elasticities. However, even in the public utility context it has proved difficult to calculate precise estimates of the relevant parameters, and the determination of optimal prices has often led to considerable controversy among economists and policymakers.

A robust conclusion of the economic literature on interchange fees and two-sided markets is that cost-based interchange fees are generally not socially optimal.[79] Even if one were convinced, as some regulators seem to be, that current interchange fees are too high, unlike the public utility case, there is no guarantee that lowering them toward any particular target will improve welfare. In particular, there is no basis for believing that any particular cost-based formula for determining interchange fees would even provide a first approximation to the socially optimal interchange fee. Nor is there any basis at the moment for believing that moving from the collectively determined interchange fee to a fee based on any formula that considers only costs would be likely to improve social welfare. Unlike the public utility situation, therefore, there is no basis in economic theory or fact for cost-based regulation of interchange fees such as the regime adopted in Australia or by the European Commission.[80]

In both Australia and Europe, the regulators (or at least their economic consultants) recognized that socially optimal interchange fees also depend on demand factors or network effects but, presumably, decided not to incorporate those factors because of the difficulty of doing so, citing instead the "objectivity" and "transparency" benefits of a cost-based measure.[81] Objectivity and transparency may have benefits, of course, but by themselves they do not, necessarily lead to greater economic efficiency.

Although cost-based interchange fee regulation could by happenstance improve the efficiency of the payment system, there are two fundamental reasons to doubt that it would regularly do so in practice. The first is that regulating the interchange fee will not necessarily have a significant effect on the variable prices paid by cardholders and therefore will not necessarily have a significant effect on the volume of transactions—generally the putative target of the intervention. The second

[79] The economic literature appears to be unanimous on this point; examples include Bergman 2005, Gans and King 2003a and 2003c (who accept for the purpose of argument the assumption that externalities between merchants and consumers are unimportant in mature card systems) and Wright 2003b.

[80] These regulatory schemes have the further peculiarity from an economic standpoint that the merchant discount charged by unitary systems is left unregulated while the merchant discount for cooperative systems is regulated through cost-based interchange fees. This favors the unitary systems and thus leads to a further economic distortion whose effects would need to be considered in evaluating the net social benefits from moving to cost-based interchange fees.

[81] Reserve Bank of Australia 2002, p. 12; Katz 2001, p. 29; European Commission 2002.

is that even if regulating the interchange fee could affect the variable prices paid by both merchants and cardholders, cost-based regulation is not capable of achieving the optimal prices except by happenstance. Whether cost-based regulation in practice would increase efficiency is unknown given the current state of theoretical and empirical knowledge.

C. Competition Policy and Interchange Fee Determination

Relatively early in the development of the payment card industry, the *NaBanco* court recognized the complexity of the role played by the interchange fee and the differences between that role and the role of an ordinary price in deciding that interchange fees should be evaluated under the rule of reason rather than be subject to the per se condemnation of price-fixing under U.S. law. We believe the recent theoretical literature supports this approach. There is a strong economic presumption that collective determination of ordinary prices harms consumers relative to uncoordinated, competitive pricing, but there is *no* economic presumption that collectively setting interchange fees reduces output or consumer welfare as compared to any other feasible regime.

Under U.S. law, to evaluate whether the pro-competitive benefits of collectively set interchange fees outweigh the anti-competitive costs (the test under the rule of reason) one would presumably have to compare collectively set interchange fees with the results of bilateral negotiations among acquirers and issuers.[82] As we have discussed, bilateral negotiation would at the very least involve high transactions costs in systems, like those in the United States, with large numbers of issuers and acquirers, and it may not be feasible in such systems. But assume it is feasible, and assume away the associated transactions costs. Then one would have to examine whether bilateral negotiations would lead to lower prices and higher output than collectively set interchange fees. Existing theory does not provide much help in predicting the outcome of such bilateral negotiations,[83] and, of course, the arguments in favor of interchange regulation imply that the output of credit card systems is already too high. Thus there is no reason at all to believe that bilateral negotiations would result in an interchange fee that would be closer to the social optimum than the collectively set interchange fee.

VI. CONCLUSIONS

Economists have only scratched the surface of the theoretical and empirical work that will be needed to understand pricing in two-sided markets in general and the determination of interchange fees in particular. Like much work in economics, many of the existing theoretical models are based on highly simplified representations of the industries in question and employ highly special

[82] That is the remedy that was sought by the plaintiff in *NaBanco*.

[83] As we noted above, Small and Wright 2000 conclude bilateral negotiation would threaten system viability, but it is unclear how robust this result is.

assumptions concerning relevant economic relationships in order to isolate individual aspects of interest. Nevertheless, several results have emerged from the literature that seem robust enough for policymakers to rely on:

1. The socially optimal prices for customer groups in multi-sided industries depend on price elasticities of demand, indirect network effects between the customer groups, marginal costs for providing goods or services to each group, and other factors.
2. Although socially optimal prices in the payment card industry depend on the same set of factors, the socially optimal interchange fee also depends on other characteristics of these industries that affect the relationship between the interchange fee and final prices. Those factors include the use of fixed and variable fees, competitive conditions among merchants, issuers, acquirers, and the nature of competition from cash, checks, and unitary payment systems.
3. Thus the socially optimal interchange fee is not in general equal to any interchange fee based on cost considerations alone.
4. One cannot presume on the basis of theory alone that the interchange fee set collectively by an association is greater than, less than, or equal to the socially optimal interchange fee.
5. One cannot presume on the basis of theory alone that movements from the collectively set interchange fee to any particular cost-based interchange fee will increase or decrease social welfare.
6. One cannot presume on the basis of theory alone that the collectively set interchange fee is greater than, less than, or equal to the interchange fee that would be set by bilateral negotiations.[84]

If regulators have concluded that multi-party payment systems have exhibited significantly sub-optimal economic performance, the current state of theoretical and empirical research leaves them with three serious challenges if they try to increase payment system performance overall and thereby raise economic welfare:

1. There is no empirical research that reliably addresses whether payment cards or any other payment mechanism is used too much or too little. Such research would need to consider the social costs and benefits of alternative payment systems and consider the effect of other market distortions.

[84] As we noted above, in the one example of bilateral interchange negotiation on which we have seen data, the Australian EFTPOS debit card system, the fee is negative, flowing from issuers to acquirers. And the Reserve Bank of Australia is considering raising EFTPOS interchange, shifting costs from consumers to merchants. See Macfarlane 2005.

2. Although it is possible that economists will be able to estimate the quantities necessary for determining optimal interchange fees, very little empirical work has been done thus far on most of the relevant quantities.
3. It is not clear that interchange fee regulation is the appropriate intervention for correcting distortions in payment systems. The interchange fee is a blunt instrument for affecting the prices faced by consumers if issuers assess fixed fees as well as variable fees.

CHAPTER

13

The Effect of Regulatory Intervention in Two-Sided Markets

Howard Chang, David S. Evans, And Daniel D. Garcia Swartz

ABSTRACT

The Reserve Bank of Australia reduced interchange fees by almost half thereby eliminating a significant source of revenue to issuers of credit cards. The purpose of this intervention was to align the prices of using various payment instruments with their social costs and thus reduce the use of cards, which the RBA viewed as a socially less efficient payment method than cash, checks, and PIN debit cards. The short-run result of this regulatory intervention has been the following: (1) Bank issuers have increased the fixed prices for cards and thereby recovered between 30 and 40 percent of the loss of interchange fee revenue; this fraction is likely to increase over time as cards renew and new solicitations go out. Bank issuers have not changed the per-transaction fees for cards much. (2) Merchants experienced a very small reduction in their costs. Both theory and limited empirical evidence suggest that the highly concentrated merchant sector in Australia has captured the reduction in interchange fees as profits and has not passed it on in the form of lower consumer prices. (3) The per-transaction price at the point of sale has not changed significantly. Merchants have not generally availed themselves of their right to surcharge card transactions and the per-transaction price faced by consumers from their card issuers has not changed much. Holding the number of cards fixed, the regulatory intervention has not altered prices in a way that could achieve the intent of the intervention. (4) There is relatively little evidence thus far that the intervention has in fact affected the volume of card transactions in Australia as intended by the regulation. (5) In the short-run, the effect of the regulation has been to transfer significant profits to the Australian merchant sector with that transfer being borne partly by bank issuers and partly by cardholders. (6) Since proprietary systems such as American Express were not subject to the pricing regulations and since American Express can enter into deals with banks to issue cards, banks have shifted volume from the regulated association systems to the unregulated proprietary systems.

I. INTRODUCTION

This chapter examines the effect of a significant regulatory-mandated alteration in pricing policy in a two-sided industry.[1] In doing so it provides empirical evidence that is helpful for understanding how two-sided industries work. It also sheds light on the interaction between the design of regulatory interventions and the pricing policies in two-sided industries.

In 2003 the Reserve Bank of Australia (RBA) mandated a reduction in the "interchange fee." In the context of a credit card transaction, this is the fee that the bank that acquires the receivable from the merchant pays to the bank that issued the card to the consumer. Three associations of banks in Australia had centrally set the interchange fee at around 0.95 percent of the transaction value. The RBA imposed cost-based regulation that resulted in a reduction of the interchange fee to around 0.55 percent. Absent any other adjustment, this 0.40 percentage-point reduction in interchange fees eliminated roughly AU$490 million in revenues that banks would have received from these fees in 2004.[2]

The credit-card industry is often cited as a classic two-sided product: it intermediates the transactions between merchants and cardholders.[3] This massive regulatory intervention therefore provides a natural experiment, almost, for assessing how an alteration in the price on one side of a two-sided industry affects the other interdependent parts of the system. In addition, more so than many interventions, the RBA's mandated reduction in interchange fees is so substantial that it provides the hope that one can determine empirically whether the intervention achieved its objectives.

After a number of studies, the RBA concluded that interchange fees that were "too high" helped sustain card transaction prices (to consumers) that were "too low" from a social perspective. This resulted in the overuse of cards relative to other, allegedly cheaper, payment instruments. The main objective of the regulation, therefore, was to raise the price of credit transactions to cardholders and thereby reduce the use of cards. Of course, at the time of this writing this experiment has only lasted for less than two years—it is possible at this time to assess only short-run effects.

In Section 2 we provide some background on the Australian credit-card industry. Then we turn in Section 3 to a summary of the RBA's theory for regulatory intervention. Section 4 provides a brief discussion of the possible effects of this intervention based on economic theory. Section 5 presents our empirical analysis and Section 6 lays out its implications for the RBA's intervention. We end with conclusions and suggestions for further research in Section 7.

---

[1] See Rochet and Tirole (2003) for the seminal work on two-sided markets, and Evans and Schmalensee (2005c) for a review of the industrial organization of two-sided markets.

[2] The reduction was 0.40 percent. According to the RBA, total nominal purchase volume on credit and charge cards was about $147 billion in 2004. Furthermore, Bankcard, Visa and MC accounted for about 83.8 percent of total volume in that year, i.e. they accounted for about $123 billion. See RBA (2005c).

[3] See Baxter (1983), Rochet and Tirole (2002, 2003) and Schmalensee (2002).

II. THE AUSTRALIAN CREDIT CARD INDUSTRY

The geography and demographics of Australia have shaped the card industry and retail sector that are the focus of this paper. Australia had a population of roughly 19 million people in 2001.[4] About 65 percent of its population lived in 8 coastal cities that year.[5] The distance between cities is great and much of the country has a very low population density.

Credit cards were first issued in Australia in 1974. Bankcard, a product that arose from the collaboration among Australian banks, was the first credit card issued in the country—by 1977 it was accepted nationally. As in the rest of the world, MasterCard and Visa marketed their products in Australia as associations of banks. Their cards started gaining traction in the mid-1980s. American Express and Diners Club came to Australia as proprietary systems, while MasterCard and Visa have had members that issue cards to consumers (issuers) and sign up and service merchants to take association cards (acquirers). American Express and Diners Club have acquired merchants on their own.[6]

Banking in Australia is relatively concentrated. There are 53 banking groups in total.[7] The four leading banks are National Australia Bank, Australia & N.Z. Banking Group, Westpac Banking Corp., and Commonwealth Bank Of Australia, which collectively account for 66 percent of total deposits.[8] Furthermore, credit card issuing and acquiring are highly concentrated. In 1999, the cards that the four major banks issued accounted for about 85 percent of all credit card transactions; these same banks accounted for about 93 percent of credit card transactions acquired.[9]

Consider the situation when a consumer presents her card for payment at a merchant. When the entity that serves the cardholder is the same as the entity that serves the merchant, the transaction is "on-us." While the entity may have transfer pricing and other matters to deal with between its issuing and acquiring "divisions," the transaction is internal to the firm. That is always the case for proprietary systems that integrate acquiring and issuing and sometimes the case for association-based systems. When the cardholder and merchant entities differ, they have to have some agreement as to who bears various risks and how the costs and benefits of the transaction get divided up. In principle that could happen through either bilateral negotiations between them, or a fee and

---

[4] See Australian Bureau of Statistics (2001).

[5] See Australian Bureau of Statistics (2003).

[6] See Reserve Bank of Australia (RBA) and Australian Competition and Consumer Commission (ACCC) (2000), p. 15. For a general introduction to payment cards see Evans and Schmalensee (2005a). American Express reached an agreement with AMP in 1998 to issue cards in Australia. More prominently, as we discuss below, major Australian banks have signed up with American Express and Diners Club since the RBA's interchange fee regulation.

[7] Excluding subsidiaries. Source: RBA, at http://www.rba.gov.au/FinancialSystemStability/FinancialInstitutionsInAustralia/the_main_types_of_financi al_institutions_in_aus.html

[8] See Australian Prudential Regulation Authority (2005). The APRA calculates total deposits as the sum of transaction deposit accounts, non-transaction deposit accounts, and certificates of deposit. It excludes intra-group deposits.

[9] See RBA and ACCC (2000), p. 17.

Table 1 Shares of major credit and charge card brands, percent of cards on issue, 1999/2000

Brand	Percent of cards on issue
Visa	51.4
MasterCard	22.7
Bankcard	19.2
American Express charge	2.8
American Express credit	2.2
Diners Club	1.7

Source: RBA and ACCC (2000), p. 15.

other contract terms that are set centrally by the association or regulators, or an industry standard that substitutes for an agreement.[10]

In Australia, Bankcard, MasterCard, and Visa adopted a multilaterally set fee.[11] Members could negotiate bilaterally if they wanted or rely on the multilateral fee. Almost all relied on the centrally set fee. As mentioned, that fee was slightly less than one percent of the transaction amount before the regulatory intervention.[12]

The interchange fee is a cost to the acquirer and is passed on to the merchant in whole or in part. The merchant discount is the percent of the transaction that the merchant pays to the acquirer. The merchant service fees for the four card brands were the following just before regulation: for Amex, 2.57 percent; for Diners Club, 2.35 percent; and for Visa and MasterCard, 1.41 percent.[13]

As of 2001, the number of merchants accepting MasterCard, Bankcard, and Visa cards was about twice as large as the number of merchant accepting American Express. The merchant base for Diners Club was apparently smaller than for American Express.[14] At this time there were roughly 13 million credit cards in use by consumers in Australia.[15] Table 1 shows the shares of each brand.

[10] See Evans and Schmalensee (2005b) for further discussion on interchange fee setting.

[11] See, for example, RBA (2001), p. 14.

[12] According to the Joint Study, the average interchange fee that issuers received in 1999 was 0.95 percent. Strictly speaking, interchange fees differed for Visa and MasterCard depending on whether the transaction was carried out with the "card present" or not. As of the end of 2001, the electronic (card-present) rate for Visa and MasterCard was 0.8 percent and the rate for all other transactions was 1.2 percent. The Bankcard banks charged 1.2 percent on all transactions, although apparently they had agreed to introduce an electronic rate of 0.8 percent toward the end of 2001. See RBA (2001), pp. 14-15.

[13] See RBA, Merchant Fees for Credit and Charge Cards, available at http://www.rba.gov.au/Statistics/Bulletin/C03hist.xls

[14] See RBA (2001), p. 119.

[15] See RBA and ACCC (2000), p. 15.

III. REGULATION AND ITS RATIONALE

The RBA and the Australian Competition and Consumer Commission (ACCC) published a "Joint Study" of payment systems in October 2000.[16] The Joint Study found that credit card interchange fees encouraged the provision of credit card services at negative prices to consumers and this fostered the use of credit cards instead of PIN debit cards, which the Joint Study believed to be a less costly instrument.[17]

In December 2001, the RBA published a "Consultation Document," outlining the need for regulation of the payment system. According to the Consultation Document, the pricing of credit card services was sending "consumers a quite misleading signal about the cost to the community of different payment instruments."[18] The Consultation Document proposed regulating the credit card schemes using "an objective, transparent and cost-based methodology for determining interchange fees."[19]

The Joint Study and the Consultation Document identified three aspects of credit card scheme rules that allegedly impeded the efficiency of the overall payments system—the collective setting of interchange fees, the "no surcharge" rule, and certain restrictions on entry to the schemes. These were the key issues addressed in the regulations introduced in the years that followed.[20] After designating the credit-card schemes as payment systems under its regulation, the RBA undertook a process of consultation and analysis to determine whether the RBA's intervention would be in the public interest. The RBA published final standards regarding interchange fees and the no-surcharge rule in August 2002 and on the access regime in February 2004.[21]

It is outside the purview of this paper to discuss the reasoning and evidence that the RBA relied on in any detail. We provide a quick summary. When customers make a purchase with their credit cards at the point of sale, the acquirer passes the interchange fee costs on to the merchant. With the no-surcharge rule, the merchant cannot effectively charge cardholders for any additional costs it incurs in accepting credit-card payments. As a result the cardholder does not have the correct incentives to use the most efficient form of payment. Social prices are not aligned with social costs and there is a resulting distortion.

The RBA buttressed this argument with evidence suggesting that credit cards are more costly than debit cards. Furthermore, the merchant passes the cost of payment methods onto its customers just like it passes all costs on. Since it cannot charge cardholders specifically for their use of cards, it passes the interchange fee—part of the merchant discount—on to all customers. The

[16] See RBA and ACCC (2000).
[17] The debit cards the RBA views as less costly are the EFTPOS PIN debit cards, not the Visa debit cards that are also offered. MasterCard does not offer a debit card in Australia.
[18] See RBA (2001), p. vi.
[19] *ibid*, p. 116.
[20] See RBA (2004), p. 7.
[21] *ibid*, p. 7.

Table 2 Timeline of RBA Investigation and Regulation

Date	Event
October 2000	Publication of the Joint Study
April 2001	Designation of Visa, MC, and Bankcard as payment systems subject to regulation
December 2001	Publication of the Consultation Document
August 2002	Publication of final standards regarding no-surcharge and interchange fees
September 2002	Visa and MC challenge standards in court
January 2003	No-surcharge standard comes into effect
July 2003	Interchange standard comes into effect
September 2003	Court rejects the Visa-MC challenge of the standards
October 2003	New interchange fees are implemented
February 2004	Publication of final standard on entry

Source: RBA and ACCC (2000), RBA (2001), and RBA (2004).

issuer, in turn, receives interchange fees as revenues. To increase this source of revenue it has incentives to encourage consumers to use their cards and it does so by providing rewards and other inducements. As a result, cardholders face negative prices for using cards and the price of using cards at the point of sale is lower than their cost. This results in a misallocation of resources (the overuse of cards and under-use of allegedly more efficient forms of payment) and a perverse distributional effect (people who do not use cards at the point of sale subsidize cardholders).[22]

In September 2002, MasterCard International and Visa International challenged the RBA's decision in the Federal Court on procedural and jurisdictional grounds. The Court rejected the challenge in September 2003, finding against MasterCard and Visa. The no-surcharge standard came into effect on January 1, 2003. The standard on interchange fees came into effect on July 1, 2003. The Bank required the credit card schemes to publish and put in force the new interchange fees by October 31, 2003. For all practical purposes, in November 2003 interchange fees declined from an average of around 0.95 percent to around 0.55 percent.[23] The timeline is shown in Table 2.

[22] We do not dispute this analysis for the purpose of this article; see Evans and Schmalensee (2005b) and Rochet (2005) for some comments. Our focus here is only on whether the RBA intervention achieved its goals.

[23] See RBA (2004), pp. 8-9.

IV. SOME THEORETICAL CONSIDERATIONS ON THE EFFECT OF THE INTERVENTION

The effect of this regulation depends on the manner in which prices are determined in this industry and the market structure of the participants. The theory of two-sided markets provides a framework for considering these issues. Businesses in two-sided markets determine price levels and price structures recognizing that the demands of their two customer groups are interdependent (Rochet and Tirole, 2003). Two-part tariffs are common in recognition of the fact that many two-sided bases involve membership and use (Evans and Schmalensee, 2005c). In the case of credit cards, card systems typically have a nominal or zero membership fee for merchants (although merchants do have to buy equipment), a membership fee for cardholders that is greater than or equal to zero (positive fees include an annual fee and various service fees), a usage fee for merchants which is usually a percent of the total transaction, and a usage fee for cardholders that is usually less than zero (float for several weeks, reward miles and other perquisites that are usually a function of volume). Credit cards bundle a transaction and finance feature; cardholders pay a finance fee on the amount they choose not to pay when their bill is due.

Average prices for reward cards with a grace period, as well as merchant fees, as of 2001, are shown in Table 3.

The regulatory intervention did not affect any of these prices directly. However, by reducing the interchange fee by almost half the intervention reduced significantly a major source of revenue to bankcard issuers. This naturally would be expected to lead them to re-equilibrate their prices. While that seems certain as a matter of economic theory, two issues remain open—the extent to which they will adjust fixed versus variable prices, and the extent to which the reduction in prices on one side will get passed on to the other side. There is, to our knowledge, no off-the-shelf theoretical guidance on how a binding ceiling on one of the four possible prices will affect the other three prices in a two-sided market. As in other markets, the extent to which the loss in revenue

Table 3 Cardholder and merchant fees, 2001

Nature of fee	Fee
I. CARDHOLDER FEES	
Annual fee – standard cards (AU$)	48
Annual fee – gold cards (AU$)	87
Late payment fee (AU$)	20
Over-limit fee (AU$)	6
II. MERCHANT FEES (*)	
Merchant service fee – Visa, MasterCard, and Bankcard	1.41 percent
Merchant service fee – American Express	2.55 percent

Sources: RBA (2005a). (*) The merchant service fees are from March 2003. The source is Merchant Fees for Credit and Charge Cards, available at http://www.rba.gov.au/Statistics/Bulletin/C03hist.xls from merchants will get passed on to cardholders depends on the degree of competition among card issuers. Given that card issuing in Australia is relatively concentrated we would not expect full pass through, at least in the short run.[24]

It is well known that in perfectly competitive markets there is full pass through of cost changes. A $1 decrease in marginal costs will lead to a $1 decrease in price. With imperfectly competitive markets the extent of pass through depends as a theoretical matter on the shape of the demand curve.[25] With linear demand curves there is less than a 100 percent pass through of costs to final consumers. For the very small cost decreases involved here it is reasonable to assume linear demand (since the curvature of demand can be ignored for such small changes) and therefore to expect less than 100 percent pass-through as a matter of theory. Empirical studies tend to find less than 100 percent pass through more frequently than greater than 100 percent pass through; the greater-than-100-percent pass through rate appears to happen in the empirical tax incidence literature and the empirical effects are confounded with the sticky-price issue discussed below.[26] Thus, we would expect less than 100 percent pass-through as an empirical matter.

---

[24] See below for a discussion of the theoretical and empirical evidence on the degree of pass through.

[25] Cotterill (1998) provides a summary of the relationship between market structure and cost pass through. If the industry is perfectly competitive and firms maximize profits, then the rate of pass through is 100 percent, no matter what the value of the market elasticity of demand is. If the industry is a monopoly, demand for its product is linear, and the monopolist maximizes profits, the pass through rate is less than 100 percent. More specifically, for a monopolist that faces a linear demand curve, the pass through rate is 50 percent. See, for example, Bulow and Pfleiderer (1983). If, on the other hand, a monopolist faces a constant-elasticity demand curve over the relevant range it is possible that pass through could be greater than 100 percent. Assuming a linear demand curve is reasonable for small changes in costs but not necessarily for larger changes in cost.

[26] The empirical literature on pass through is fairly vast and covers a number of areas, including exchange rate and tax rate pass through to prices. The literature on pass through of exchange rate changes to domestic prices of traded goods finds that the median rate of pass through is roughly 50 percent for shipments to the US. See Goldberg and Knetter (1997). In the context of the proposed merger between Staples and Office Depot, Ashenfelter et al (1998) made a distinction between the firm-specific cost pass through rate and the industry-wide pass through rate—i.e., the reaction of the firm's prices to changes in its own costs and to changes in costs common to all firms in the industry. They found that the firm-specific pass through rate for Staples was roughly 15 percent, whereas the industry wide pass through rate was around 57 percent. Furthermore, there is a literature focused on pass through of trade promotions to retail prices, which is a form of firm-specific pass through. There are a number of relevant studies in this area, including Chevalier and Curhan (1976), Walters (1989), Armstrong (1991), and Besanko, Dube, and Gupta (2005). Besanko et al (2005) analyze the degree of pass through of trade promotions using scanner data for eleven product categories at a Chicago supermarket chain. 8 out of 11 categories show pass through rates of less than 100 percent. At a more disaggregated level, only 165 products out of 1164 (about 14 percent) show pass through rates significantly larger than 100 percent. Besanko et al (2005) cite a study of trade promotions according to which retailers themselves said that they pass through roughly 62 percent of the promotion, use 24 percent to cover promotion costs, and keep the rest as profits. In the context of this literature, Blattberg et al (1985) suggest that, as an empirical matter, most products display pass through rates much smaller than 100 percent. Furthermore, Tyagi (1999) suggests that pass through rates of less than 100 percent should occur much more frequently than pass through rates of more than 100 percent, since the set of demand functional forms that imply less than 100 percent pass through is quite large. It seems to us that pass through rates of more than 100 percent are less frequent as an empirical matter than rates of less than 100 percent, and they tend to appear most often in the tax incidence literature. Poterba (1996), for example, is unable to reject the hypothesis of 100 percent pass through for the commodities he examines, but Besley and Rosen (1999) find greater than 100 percent pass through for several of the commodities in their sample. In line with our discussion of sticky prices below, note that empirical estimates of pass through mix pass through effects and sticky price effects.

There is a further consideration, from a theoretical perspective, on how the regulatory intervention would affect final good prices. It is well known that many prices are sticky in the short run (Stigler and Kindahl, 1970; Carlton, 1986). Of particular interest in our case is the finding that prices tend to rise faster than they fall. A number of studies in the 1990s found that retail prices respond faster to input price increases than to input price decreases (Karrenbrock, 1991; Newmark and Sharpe, 1992; Borenstein, Cameron, and Gilbert, 1997; Jackson, 1997). More recently, Peltzman (2000) confirmed this finding with a large sample of consumer and producer goods. This is especially important for considering short-run versus long-run effects of the intervention.

The credit-card acquiring business in Australia is highly concentrated. The four major banks accounted for 93 percent of credit card transactions acquired in 1999.[27] In contrast, the share of the four major acquirers in the United States is around 41 percent.[28] As we note below, the RBA has reported that the reductions in the interchange fee appear to have been fully passed through to merchants. This is not surprising because, despite the relative concentration of acquirers, many merchants are large customers and can bargain effectively with acquirers. In addition, the acquirers are aware that their actions have been monitored closely by the RBA. As noted above, credit-issuing is also highly concentrated. We would therefore not anticipate full pass-through and, at least for the short-run, that is what we find.

Comprehensive government statistics on merchant concentration are not available for Australia. We have identified what data we could find from a range of sources. Many merchant categories appear to have significant levels of concentration. For example, within their respective categories, the top department store had a 71 percent share (and the top two had an 83 percent share) in 2003;[29] the top three supermarket and grocery stores had a 75.4 percent share in the late 1990s;[30] the top two mobile telephone operators had a 78 percent share in 2005; the top land-line telephone operator had a 75 percent share in 2005;[31] and the top two airlines had an 83 percent share in 2005.[32]

---

27 See RBA and ACCC (2000), pp. 16-17.

28 These calculations are based on data from HSN Consultants Inc. (2003).

29 See Euromonitor International (2005a).

30 See http://www.aph.gov.au/senate/committee/retail_ctte/report/c04.htm. The Australian Bureau of Statistics reported the 75.4-percent figure, which was based on sales of all supermarkets and grocery stores, including the non-petrol sales of convenience stores at petrol stations. Furthermore, Euromonitor International (2005b) estimated that, in 2003, the size of the total grocery stores/food retailers/supermarkets sector was AU$ 72.5 billion and the supermarket sub-sector represented 60.5 percent of that. The top two supermarkets had a combined market share (in the total sector) of 37.9 percent. Therefore, they had a combined market share in the supermarket sub-sector of 62.6 percent.

31 See Maxwell (2005).

32 See Bureau of Transportation Statistics (2000).

There are significant differences between the cardholder and merchant sides of the business that likely affect relative pass through rates to consumers on each side. The regulatory intervention caused issuers to experience a significant reduction in revenue (which could also be viewed as an increase in costs). That makes menu costs and other sources of sticky prices less binding. Moreover, it appears that it is quite easy for issuers to adjust prices by varying service and other card fees. For example, it seems that average annual fees on standard and gold cards changed (in a non-trivial manner) every year between 2000 and 2004.[33] Lastly, linear demand is a more reasonable assumption for the merchants who experience a small relative cost decrease than for issuers who experience a relatively large cost increase (decreased interchange fees).

The main source of friction for issuers concerns the annual fee and other prices that issuers commit to in trying to persuade people to take their cards or to switch from another card. Cards are replaced about every three years; consequently that is the opportunity to institute or increase annual fees for current cardholders. For new solicitations it is easy to change fees whenever the solicitation goes out.

Merchants on the other hand experienced a relatively small reduction in cost. If fully passed through by acquirers, the interchange fee reduction amounts to less than half a percent of their selling price (and only on those transactions that take place on credit cards). The evidence on price rigidities, and particularly the one on asymmetric price responses cited above, makes one doubt that such a small cost reduction would affect final goods prices very quickly, even if there were extensive retail competition. We return to this point below.

V. EMPIRICAL ANALYSIS: PRICES AND QUANTITIES

Although there was a sharp reduction in interchange fees on a particular date as a result of the RBA intervention, we cannot simply compare markets before and after this intervention. The credit-card industry learned over the course of several years that an intervention was increasingly likely. Furthermore, over time it developed a sense of the impact that the intervention would have. We thus begin by examining the time line a bit more carefully.

Although regulated interchange fees did not come into effect until the end of October 2003, there was anticipation that regulation was coming. That is not to say, however, that even if a bank in say January 2002 was certain that regulation would be implemented in October 2003, it would necessarily have changed its pricing or strategy in January 2002, since interchange fee levels for itself and its competitors were still at prior levels. Nor does it mean that a bank would necessarily wait until the actual date of the regulation to change its behavior. A bank would not offer a cardholder annual fees in, say, September 2003 based on pre-regulation interchange fee levels if it knew that the post-regulation levels would be in place two months later.

[33] See RBA (2005a).

Our best estimate on timing is that there were at least some changes in bank behavior in early 2003 in anticipation of the regulatory intervention.[34] The annual report for one of the major banks, ANZ, noted that in 2003 it had reshaped its "product set across the Australian Cards Issuing portfolio to address the impact of the Reserve Bank interchange reforms."[35] There are a number of newspaper articles along the same lines in the period prior to October 2003.[36] The *Sunday Telegraph* (Sydney), for example, reported in its September 21, 2003 edition that the five major banks had increased credit card fees by up to 50 percent in the previous 12 months. The article goes on to quote a bank executive who expressed that the rise in fees had the goal of making up for the loss in interchange income that would happen after the implementation of the regulatory scheme.[37] In our analyses below, we consider whether we see any effect of regulation starting in 2003 and, alternatively, starting in 2004.

We analyse two questions. (1) How did the intervention affect prices to issuers, cardholders, merchants, and consumers? (2) What was the effect of the intervention on card use? Appendix A provides information on our data sources and detailed statistical results for interested readers.

A. Effect on Prices

1. Issuers and Cardholders

Visa Australia provided the data used in the calculations in this section. The dataset was constructed with information from the operating certificates that banks submit to the Visa organization. The dataset provides quarterly information on the number of Visa credit cards, credit card purchase volume, other service charges (i.e., fees that issuers charge cardholders, primarily annual fees and service fees), finance charges, and outstanding balances on credit cards in Australia between the third quarter of 1992 and the first quarter of 2005.[38]

Between the last quarter of 1992 and the fourth quarter of 1999, real interchange income per Visa card grew at an average quarterly compound rate of about 4.2 percent, from around AU$5.79 to around AU$18.34, driven by the rise in purchase volume per card.[39] Between the first quarter of 2000 and the third quarter of 2003, the quarter prior to the introduction of the new interchange rates, real interchange income per card grew at an average quarterly compound rate of about 3.12 percent, from around AU$17.26 to around AU$26.55.

---

[34] Some changes in fees may have been happened already in 2002.
[35] See ANZ (2003), p. 31.
[36] See, for example, McKinnon (2001), Hanna (2002), Brammall (2002, 2003), Horan (2003), and Graeme (2003).
[37] See Horan (2003).
[38] At the time of this writing the information on other service charges and finance charges was available only through the second quarter of 2004.
[39] As noted above, all figures are reported in real 2004 Australian dollars.

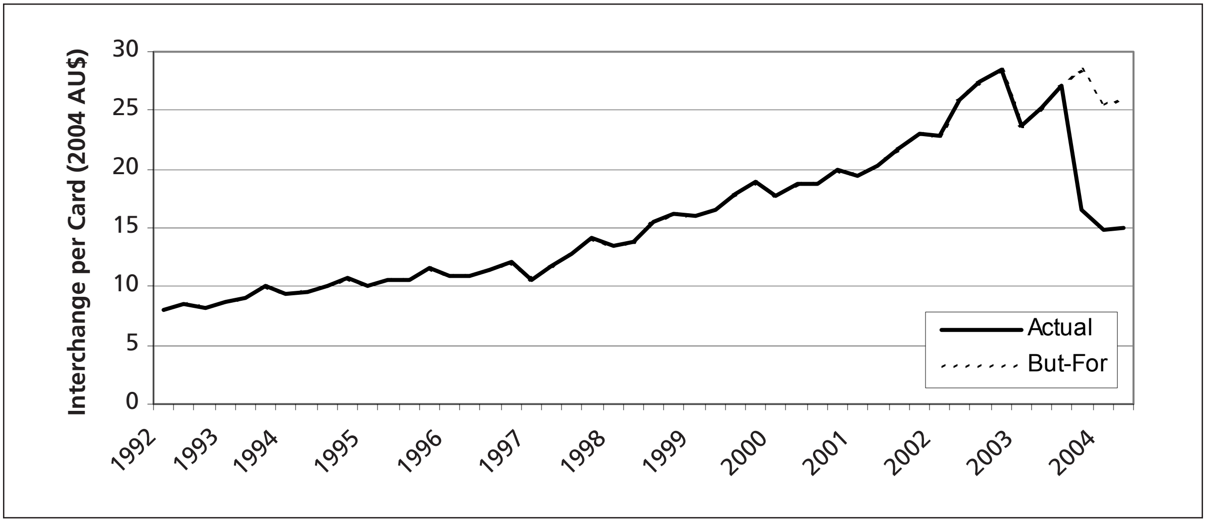
Interchange-fee regulation was implemented in the fourth quarter of 2003. If we compare the first two quarters of 2003 with the first two of 2004, we find that issuers lost, on average and in real terms, about AU$9.35 per card in interchange income per quarter, a loss of about 40 percent on total interchange income per card of about AU$23.52 per quarter in the first two quarters of 2003.

A more interesting way to look at this loss is to ask how much the issuers would have made in interchange income in the absence of regulation. From this perspective, we calculate a "but for" interchange income per card under the assumption that the regulatory scheme was not implemented and the interchange rate remained at an average of 0.95 percent. We then subtract the actual interchange income per card from the "but for" interchange income per card and we find that, on average, in the first two quarters of 2004 issuers lost about AU$10.31 per quarter per card. This represents a loss of about 42 percent with respect to the interchange income they would have obtained in the "but for" world. Figure 1 shows these facts.

The evidence thus suggests that issuers started losing, in real terms, roughly between AU$9.30 and AU$10.30 per card per quarter in interchange income as a consequence of the RBA intervention. In the months that followed the introduction of the regulation (and likely in the months that preceded the regulation as well) they recovered between 30 and 40 percent of that loss through the imposition of higher fees.

Appendix A provides the support for this conclusion. A regression model that compares the level of real other service charges per card after regulation with the level before, controlling for seasonal effects, changes in the real purchase volume per card, and changes in unobserved factors over time, suggests that the regulation (or its anticipation) was accompanied by a rise in real other service charges of roughly 30 to 40 percent of the interchange loss amount. (As we

Figure 1 Real Interchange Revenue per Card



discuss further in the appendix, the quarterly Visa data we used track closely with annual RBA data on fees. The increase in issuer annual and service fee revenue per card from 2001 to 2003 was AU$20.84 from the RBA data and AU$22.78 from the Visa data.) A similar model for real finance charges per card produces positive coefficients for the post-regulation period. The coefficients, however, are for the most part not statistically significant. Furthermore, we found no evidence of a structural break (associated with the RBA regulation) in the interest rate that issuers charge on outstanding balances.

The components of the "other service fees" variable from the Visa data are annual fees and service fees. Annual fees are fixed with respect to transaction volume. Service fees (such as late payment fees and over-limit fees) are also primarily fixed fees. A late payment fee, for example, is independent of the amount charged.[40]

In addition to the increases estimated above for these fixed fees, there has also been an impact on the level of rewards offered by issuers. The RBA has reported that the average reward decreased from 0.8 to 0.6 percent for most of the bank schemes.[41] While rewards (miles or points accumulated as a function of purchase volume) affect marginal incentives to use cards, the way in which issuers have implemented changes has likely limited the impact on purchase volume. First, some of the changes have been in the form of caps on the number of reward points accumulated. For example, National Australia Bank (NAB) capped rewards on its Visa rewards card. It offers 1 point for each dollar spent up to AU$3,000 per month. Above the $3,000 a month spending threshold, cardholders receive 1 point for every two dollars spent, up to a maximum of 13,000 points a month. For cardholders below the AU$3,000 threshold, the marginal incentive to use cards has not changed as a result of the imposition of the cap.[42]

For customers who are likely to exceed the monthly cap on a regular basis, the marginal incentives have changed on the NAB Visa rewards card. However, one of NAB's other responses to the regulation was to partner with American Express, in offering a rewards card without a cap (in fact offering 1.5 points for each dollar in the first year). Similarly, Westpac now offers an American Express rewards card and ANZ offers a Diners Club rewards card, both without rewards caps and

---

[40] The over-limit fee may be partially variable, in the sense that a consumer may be more likely to exceed the credit limit at higher purchase volumes. But it is not clear that higher over-limit fees would significantly affect consumer purchase volume on credit cards. Among other things, over-limit fees would appear to be more directly related to accumulated revolving balances as opposed to purchase volume (although the two are related).

[41] See Testimony of Dr. Philip Lowe, House of Representatives Standing Committee on Economics, Finance and Public Administration (EFPA) (2005), at p. 26-27. There is insufficient detail on this data point for us to conduct a detailed analysis. The RBA does not report, for example, how this estimate is constructed, nor does it provide a data series over time (or even which years are the beginning and end points). It is also unclear whether this includes rewards for cards issued on the American Express and Diners Club systems.

[42] The cost of having rewards cards has increased, which may result in fewer consumers deciding to hold them, but contingent on having a card, the marginal incentive to use a card has not changed for consumers below the threshold.

both more generous than the MasterCard/Visa rewards cards from the respective banks.[43] As we note below, while the interchange fee regulation appears to have had relatively little effect to date on overall card volume, it does appear to have switched cards from MasterCard and Visa to the American Express and Diners Club systems.[44]

What we have estimated is the short-run impact. Since cardholders are valuable assets—issuers spend considerable resources to acquire new cardholders and pay a premium on portfolio purchases—one would expect issuers to be cautious in raising prices quickly to their existing mature cardholder portfolio. Over the longer run, as cardholders switch from one issuer to another, one would expect price effects to be more fully realized.

2. Acquirers, Merchants, and Consumers

According to the RBA, the reduction in interchange fees imposed by regulation led to a reduction in merchant service fees (i.e., the merchant discount). The average merchant service fee that the now regulated systems charge fell from 1.41 percent immediately before regulation to 0.99 percent in the quarter ending June 2004.[45] The RBA estimates that the fall in the merchant discount represents savings to merchants of over $500 million per annum.[46]

We consider the extent to which this decrease in costs to merchants was passed on to consumers.[47] To begin with, the reduction in cost was quite small for retailers. The cost of a credit-card transaction made with a BankCard, MasterCard or Visa card fell by 0.42 percent. However, these transactions comprise only about a quarter of retail transactions.

According to the RBA, total purchase volume on credit and charge cards in 2004 was roughly AU$147 billion, and the regulated systems accounted for 83.8 percent of that (i.e., roughly 123 billion). Total household consumption in 2004 was roughly AU$500 billion, so that purchase volume on regulated credit cards represented roughly 25 percent of total consumption. To be con-

[43] See, Westpac Bank, at http://www.westpac.com.au/internet/publish.nsf/Content/PBCCCSCR+Altitude; ANZ Bank, at http://www.anz.com/australia/support/library/MediaRelease/MR20030912.pdf

[44] See EFPA (2005), p. 25. See also, Diners Club of Australia website, at http://www.dinersclub.com.au/s06_media/p62_view.asp?id=144

[45] See RBA (2004), p. 9.

[46] We have not made any attempts to verify this independently. In any case, according to the RBA, total nominal purchase volume on credit and charge cards was about $147 billion in 2004. According to the RBA, Bankcard, Visa and MC accounted for about 83.8 percent of total volume in that year, i.e. they accounted for about $123 billion. The reduction in the merchant discount was 0.0042 percentage points, which gives savings to merchants of roughly $517 million.

[47] The RBA claimed in testimony before the Australian House of Representatives in 2005 that these cost savings would be passed on almost entirely to final consumers. It reasoned that the retail sector overall was not concentrated even though particular segments such as supermarkets were. This analysis ignores the point, of course, that the extent of pass through depends on the structure of the relevant market, which is not overall retail but the individual markets that comprise it. Testimony of Mr. Ian MacFarlane, House of Representatives Standing Committee on Economics, Finance and Public Administration (EFPA) (2005), at pp. 23-24.

servative, we take the card share of retail transactions to be 50 percent, which will likely overstate the card share for many merchants, especially those in the service sector.[48] Taking this 50 percent estimate, then the average reduction in overall merchant costs as a result of the interchange fee reduction was 50 percent of 0.42 percent or 0.21 percent.

There are three reasons, mentioned earlier, why we would expect that prices to consumers would have fallen by less than 0.21 percent. First, this decrease in costs amounts to roughly 8 cents on an AU$40 transaction. The empirical literature on price rigidities makes it doubtful that a decrease in cost of this small magnitude would be passed on to consumers quickly. Second, many of the significant retail markets in Australia are highly concentrated. Given the small decrease in cost it is reasonable to approximate the demand curve facing these merchants with a linear demand curve. In this case, pass through rates will be less than 100 percent. Third, it seems that the empirical evidence on pass through—while not specific to Australia—finds rates of less than 100 percent more often than rates of more than 100 percent.[49] With a 50 percent pass through rate, the reduction in prices to consumers from the RBA's interchange fee reduction would be 0.105 percent. That amounts to roughly 4 cents on a AU$40 transaction. It would not require much in the way of price rigidities for merchants to decide not to adjust prices in the short run.

The very little empirical evidence there is suggests that, in fact, merchants have tended not to pass through the reduction in the merchant discount to consumers in the form of lower prices. Cannex, an independent research group, surveyed merchants in Australia regarding the impact of the interchange fee regulation on their regular business practices.[50] Among merchants who reported a change in the merchant discount during the previous year, less than 5 percent declared that they had reduced prices to consumers. On the other hand, more than 20 percent reported that their profits had increased and almost 60 percent reported that they had not experienced any changes in their regular operations. Surcharging for credit card transactions, following the repeal of non-surcharge rules, is still the exception in Australia, as it has been in other countries.[51] One survey of Australian merchants in November 2004 found that only 2.3 percent of all merchants surcharged, with larger merchants slightly more likely to surcharge, at slightly over 5 percent.[52] A

48 The 25 percent figure of share of cards of total consumption may be an underestimate of the percent of retail transaction dollars on cards since the consumption figure includes some non-retail merchants where cards are not commonly used. Unfortunately, more precise data are not available for Australia. In the United States, where the card share of total consumption is only slightly higher than Australia, the portion of transactions at merchants paid for with cards is slightly under 50 percent for retail and travel and entertainment merchants, and less than 10 percent for service merchants.

49 See *supra* notes 25-26.

50 See Cannex Australia (2004).

51 Only about 10 percent of merchants imposed surcharges in the Netherlands, and only about 5 percent of merchants imposed surcharges in Sweden. Surcharging is also uncommon in the United Kingdom. See Evans & Schmalensee (2005b), p.27.

52 See Network Economics Consulting Group (NECG) (2005), p. 46, citing East & Partners 2004, "Australian merchant acquiring and cards markets: Multiclient market analysis report," December.

different survey found that 7 percent of all merchants surcharged regularly.[53] It is also worth noting that the average surcharge was 1.8 percent, which is higher than the merchant discount fee on credit card transactions, and also almost surely higher than any differential between the costs to merchants of processing credit cards versus other forms of payment. This suggests that at least some of the surcharging that takes place may be opportunistic and does not increase the efficiency of relative prices for payment instruments facing consumers.

B. Effect on Transaction Volume

If the preceding estimates are correct we can make some surmises about the effect of the regulatory intervention on transaction volumes. It appears that cardholders are not facing substantially different prices at the point of sale for using credit cards. The usage prices assessed by the issuers do not appear to have risen generally and remain negative in many cases. At the same time the preponderance of merchants are not availing themselves of surcharging. Therefore, holding the number of cards constant we have no reason to expect more than a modest change in the volume of transactions in the short run.

However, the increase in fixed fees means that we would expect that fewer individuals have cards. The elasticity of demand of card membership with respect to membership and usage fees determines the relative decline in cardholders. We would expect that any decline in usage of MasterCard and Visa cards as a result of a decline in membership would take place gradually as people adjust the number of cards they have. A further complication is that the RBA did not impose any price regulation on American Express and Diners Club. Some banks have started issuing American Express cards, in particular, in response to the higher effective price they receive for those cards relative to the regulated MasterCard and Visa cards. Therefore, from the standpoint of a regulatory desire to reduce the use of cards, we need to consider total credit-card transactions and not just those of the regulated systems.

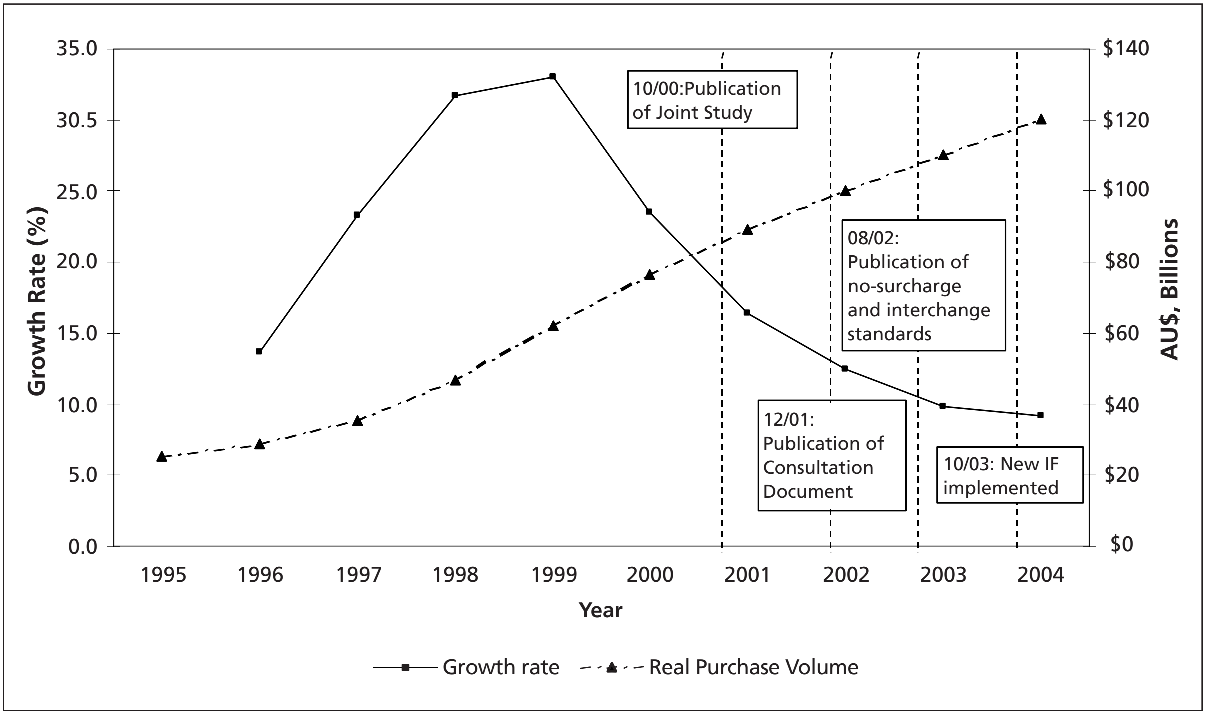
Despite a rather massive regulatory intervention that eliminated about 30 percent of issuer revenue in the stroke of a pen,[54] there is little evidence to date that the intervention has significantly affected the use of credit cards in Australia. At the same time it appears that some proportion of transaction volume has moved from association cards to proprietary cards.[55]

[53] The survey reported that 19 percent of merchants "sometimes" surcharged, but it is unclear how commonly and on what criteria they did so. Two other surveys that asked incidental questions on surcharging reported that 12 percent and 2.9 percent, respectively, of respondents surcharged (although the frequency of surcharging was unknown). See NECG (2005), at p.44.

[54] In this calculation we are excluding revenues derived from the so-called "interest margin."

[55] According to the RBA data, the regulated systems accounted for 86.3 percent of the total value of credit card purchases in September 2003. In June 2005 they accounted for 83 percent. See the market-share data available at http://www.rba.gov.au/Statistics/Bulletin/C02hist.xls

FIGURE 2 Real Purchase Volume on Credit Cards: 1995 – 2004



Source: RBA (2005b).

Figure 2 shows the level and growth rate of real purchase volume on credit cards along with key points in the timeline for the regulatory intervention. Despite the massive degree of the RBA's intervention, there appears to be no evidence of any effect of the intervention on the use of cards. Table 4 goes into more detail by showing the trends over time in several measures of card use.

Other than the number of accounts, the indicators of credit-card use grew at an increasing rate through 1999 and then grew at a declining rate through 2002. Assessing the impact of the RBA's regulation depends on our assumptions about card industry performance in the absence of regulation. For example, if the question is whether the regulations stopped the growth of card ownership and usage, then the answer is "no"—all card metrics continued to grow post-regulation. Our main focus is on two different questions, namely (1) whether growth rates were lower post-regulation than pre-regulation and (2) whether the decline in the growth rates that was taking place pre-regulation accelerated or decelerated.

The aggregate data suggest that (1) industry growth was lower after regulation than before, with the exception of the number of accounts, and (2) the decline in the rates of growth that started around 1999 has continued, again with the exception of the number of accounts, but has not accelerated. These basic empirical regularities are confirmed by our detailed analysis discussed below. The first comparison would say that regulation has lowered the growth rates, while the second would say that the decline in the growth rates was already taking place and the regulation had

Table 4 Annual Growth Rates of Indicators of Credit Card Use, All Credit Cards, 1996-2004

	Accounts	Purchase Transactions	Purchase Transactions per Account	Real Purchase Volume	Real Purchase Volume per Account
1996	7.7	14.2	6.1	13.7	5.6
1997	5.6	19.7	13.3	23.3	16.7
1998	6.1	30.8	23.2	31.7	24.2
1999	5.9	32.5	25.1	33.0	25.6
2000	7.0	24.0	15.9	23.5	15.5
2001	2.5	16.9	14.0	16.4	13.5
2002	1.5	13.9	12.2	12.4	10.7
2003	4.7	8.9	4.0	9.9	4.9
2004	6.0	8.7	2.5	9.2	3.0

Source: all the growth rates have been calculated on the basis of RBA (2005c).

relatively little impact. Our prior is that the second comparison is more appropriate, as we observe a significant slowdown in the rate of growth in the pre-regulation period. We acknowledge, however, that our analysis does not come close to explaining the dramatic shifts in industry output that have taken place— either the accelerating growth leading up to 1999 or the decelerating growth following 1999. Therefore, we believe it is at least possible that absent regulation, growth rates of the different card metrics might have levelled off, or even increased.[56]

In order to address the first question—i.e., whether growth rates were lower post-regulation—we estimated a number of simple regression models of the quarterly growth rates of the credit card use variables on a set of quarterly binary variables, a binary variable that takes on the value 1 between 2000 and 2002 and 0 otherwise,[57] and a binary variable that takes on the value 1 starting in 2003. The results are reported in Table 4a.[58]

[56] For example, the entry of Virgin as an issuer in 2004 could have increased output, all else equal. To the extent that is true, our comparisons would underestimate the impact of the RBA's regulation. We do not have detailed and systematic data on industry concentration, or on entry and exit, to attempt to control for these factors.

[57] In the accounts model this binary variable takes on the value 1 in 2001-2002 rather than in 2000-2002.

[58] All the regressions include binary variables for the second, third, and fourth quarters on the right hand side. When the growth rate in the number of accounts is the dependent variable, we control for changes in population and income per capita. In the case of the number of purchases and real purchase volume, we control for changes in real consumption. In the case of the number of purchases per account and real purchase volume per account, we control for changes in consumption per capita.

Table 4a Regression of the quarterly growth rates of the credit-card-use variables on a binary variable for 2000-2002 and a binary variable for the regulated period (P-values calculated on the basis of Newey-West standard errors)

	Coefficient	P-value
Growth in Number of Accounts		
2001–2002	20.0096	0.000
2003–2005	20.0013	0.682
P-value for F test of equality between binary-variable coefficients = 0.0116		
Growth in Number of Purchases		
2000–2002	−0.013	0.214
2003–2005	−0.033	0.001
P-value for F test of equality of binary-variable coefficients = 0.0184		
Growth in Number of Purchases per Account		
2000–2002	−0.007	0.362
2003–2005	−0.032	0.002
P-value for F test of equality of binary-variable coefficients = 0.0026		
Growth in Real Purchase Volume		
2000–2002	−0.012	0.278
2003–2005	−0.033	0.001
P-value for F test of equality of binary-variable coefficients = 0.0259		
Growth in Real Purchase Volume per Account		
2000–2002	−0.007	0.445
2003–2005	−0.032	0.002
P-value for F test of equality of binary-variable coefficients = 0.0021		

The interpretation of these results is straightforward. In the case of the number of accounts, growth rates in the regulated period have been no different from what they were through 2000 and higher than they were in 2001-2002. In the case of all other variables, however, growth rates in the regulated period have been lower than they were through 1999 and also lower than they were in 2000-2002. From this perspective, the regulated period has been associated mostly with lower growth rates in the indicators of credit card use.

We then addressed the second question—i.e., has the decline in the growth rates that started in the late 1990s accelerated or decelerated with regulation? We used regression methods to examine more carefully whether credit card activity grew more slowly than it would have in a world without the RBA regulation. We studied the growth patterns in greater detail by regressing the growth rates of the relevant variables on a linear spline function.[59] We performed the analysis with annual data and quarterly data. With quarterly data, we calculated proportional growth rates between quarter q and quarter *(q-1)*.[60]

We describe the model on the basis of quarterly data. We first created a linear trend t that starts at 1 in the first quarter for which growth rates can be calculated and grows by 1 every quarter through the first quarter of 2005. We then defined the spline time dimension variables as follows:

$$X(t) = t,\ t = 1, 2, \ldots, T;$$
$$Y(t) = \max(0, t - a);$$
$$Y(t)' = \max(0, t - b)\text{ ; and}$$
$$Z(t) = \max(0, t - c).$$

In this model, *a* is the number that corresponds to the last quarter of 1999 in the trend sequence, *b* is the number that corresponds to the last quarter of 2000, and c is the number that corresponds to the last quarter of 2002.

Think of a world where the number of credit card accounts, the number of credit card purchases, and the real purchase volume on credit cards are functions of a set of demand shifters. For example, the number of credit card accounts is a function of population and income per capita, and the number of credit card purchases is a function of real consumption. Then in order to assess whether there have been significant changes in the trend of the growth rates of the credit card use variables, we can estimate models of the following form:

$$G_{CC}(t) = \alpha_1 + \delta_{11} X(t) + \delta_{12} Y(t) + \delta_{13} Z(t) + \gamma_1 C(t) + u_1(t), \text{ or}$$
$$G_{CC}(t) = \alpha_2 + \delta_{21} X(t) + \delta_{22} Y(t)' + \delta_{23} Z(t) + \gamma_2 C(t) + u_2(t).$$

In these models, $G_{CC}(t)$ stands for the growth rate of the credit-card-use variable in question and C(t) stands for the growth rate of the relevant demand shifter. These models focus on the percentage-point change in growth rates over time.[61]

---

[59] On linear spline functions see, for example, Poirier (1976), ch. 2; Poirier and Garber (1974); Johnston (1983), p. 392ff; and Greene (1993), p. 235ff. For a study that uses linear splines in the context of testing for the existence of unit roots in economic time series, see Perron (1989). For a discussion of the various Perron models, see, for example, Enders (2004), pp. 200-207.

[60] We also performed the analysis with the growth rates calculated between quarters *q* and *(q-4)*. Although the magnitude of the coefficients was, of course, different, the substance of the conclusions we draw did not change much.

[61] The results we obtain do not change in any substantive manner if we exclude the demand shifters from the estimated regressions.

The δ_{i1} parameters (for i = 1, 2) capture the trend in the growth rate between the starting point and 1999 (or between the starting point and 2000). The δ_{i2} parameters capture the change in the trend of the growth rate over 1999-2002 relative to the previous period (or 2000-2002 relative to the previous period). The δ_{i3} parameters, finally, capture the change in the trend of the growth rate over 2003-2004 (the regulatory period) relative to 1999–2002 (or 2000–2002).

The estimated parameters with the p-values calculated on the basis of the Newey-West standard errors are reported below in Table 5.[62]

The results, based on quarterly growth rates, show the following. All of the models estimated show a positive point estimate for the trend that starts in 2003. In three of the models, the relevant coefficients are statistically significant—the exceptions are purchases per account and real purchase volume per account. These models convey the idea either that the trend of growth of the relevant card-use variable accelerated during the regulated period (e.g., number of accounts) or that the decline in the trend of growth that had started in the late 1990s decelerated during the regulated period (e.g., number of purchases).

The two models that raise some questions are the ones estimated in the growth rate for the number of purchases per account and real purchase volume per account. They both show a positive coefficient for the trend that starts in 2003 but the coefficients are not statistically significant. (A simple examination of Table 4, furthermore, shows that these two series seem to have exhibited considerably lower rates of growth in the post-regulation period than in the pre-regulation period as compared to other indicators of credit card use.) Whether the regulation had any impact in terms of curtailing purchases per account and purchase volume per account is thus an open issue.

Overall, taking into account existing trends in the growth rates of card variables, there is little evidence that the regulatory intervention has affected overall card use in the admittedly short period of time examined here. Figure 3 below, however, reveals that there have been important compositional changes—volume has shifted from the regulated systems to the unregulated ones.[63] More specifically, between October 2003 and June 2005 the share of American Express and Diners

[62] All the regressions include binary variables for the second, third, and fourth quarters on the right hand side. When the growth rate in the number of accounts is the dependent variable, we control for changes in population and income per capita. In the case of the number of purchases and real purchase volume, we control for changes in real consumption. In the case of the number of purchases per account and real purchase volume per account, we control for changes in consumption per capita. We also tried estimating the models via instrumental variables—we instrumented the growth in real consumption and the growth in real consumption per capita. We calculated the Newey-West standard errors for the IV-estimated coefficients. Overall, as far as the trend coefficients are concerned, the results were not substantially different from the ones obtained via OLS with Newey-West standard errors. We also calculated feasible-generalized-least-squares (Prais-Winsten) estimates. The conclusions we draw did not change under the FGLS approach.

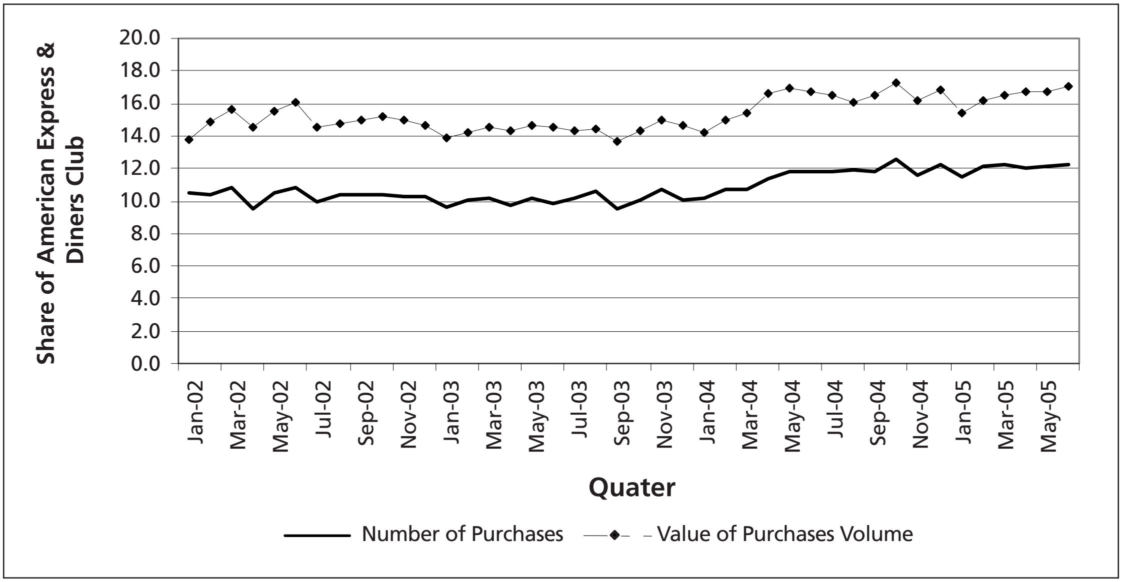
[63] Analysis of Visa data on card and volume also indicates that Visa output declined relative to total industry output following regulation.

Table 5 Regression of the growth rate of the credit-card-use variables on a linear spline function on the basis of quarterly data, 1994-2005 (P-values calculated on the basis of Newey-West standard errors)

	Coefficient	P-value
Growth in Number of Accounts		
d(1)	0.0001	0.592
d(2)	−0.0018	0.001
d(3)	0.0039	0.000
Growth in Number of Purchases		
d(1)	0.0021	0.000
d(2)	−0.0066	0.000
d(3)	0.0044	0.003
Growth in Number of Purchases per Account		
d(1)	0.0018	0.001
d(2)	−0.0049	0.000
d(3)	0.0010	0.499
Growth in Real Purchase Volume		
d(1)	0.0020	0.002
d(2)	−0.0063	0.000
d(3)	0.0044	0.005
Growth in Real Purchase Volume per Account		
d(1)	0.0016	0.014
d(2)	−0.0047	0.001
d(3)	0.0010	0.509

Club increased by roughly 21 percent in terms of the number of purchases and by roughly 19 percent in terms of the value of purchases. This has happened because, as a consequence of regulation, relative prices seem to have changed. First, the relative price that issuers receive from a Visa or MasterCard transaction has declined relative to, say, an American Express transaction, since interchange has been capped for Visa and Master Card but American Express is allowed to sign issuing deals with banks under no interchange ceiling. Therefore, issuers have an incentive to shift volume

Figure 3 Share of American Express and Diners Club (in terms of number of purchases and value of purchases)



Source: RBA (2005d)

to American Express, and they have. Secondly, the price of a card (or account) that cardholders face may have increased for the Visa and MasterCard cards relative to the proprietary ones, since Visa and MC issuers have attempted to recover some of the lost interchange income via an increase in other service charges, most of which are fixed fees. (Of course, in light of the fact that Visa and MasterCard issuers have raised other service charges, American Express and Diners Club may have done the same, so it is not altogether clear whether this relative price has changed much or not.)

Note, by the way, that the shift in volume from the regulated systems to the unregulated ones has a "perverse" effect on prices. In other words, the merchant discount on the regulated systems has come down significantly and the merchant discount on the unregulated systems has declined slightly, but it is considerably higher than the associations' pre-regulation merchant discount. Although the pure price effect leads to a lower average merchant discount, the compositional-change effect leads to a higher average merchant discount. (Of course, it will take an extremely large compositional change effect to produce an average merchant discount rate that is higher than the average pre-regulation rate.)

VI. ANALYSIS OF THE RBA INTERVENTION

The RBA's regulation of interchange fees has had some economic consequences that were entirely predictable and hardly surprising and others that raise some interesting questions both for regulators and students of two-sided markets.

One predictable result is that a massive reduction in revenue from one side of a two- sided market had consequences on the other side. Banks lost roughly $490 million in interchange fee revenue but appear to have regained between 30 and 40 percent of that through increased membership fees for cardholders. We believe this is a lower bound on the portion passed on to cardholders because of the cycle of replacing cards.

Another predictable result is the absence of evidence that consumer prices have fallen as a result of lower merchant discounts. This is not surprising because the cost savings are too small to be measurable with any degree of confidence. However, based on the economics literature on pass-through effects, we believe that it is highly unlikely that consumers have received any significant benefit over the period of time considered given the likely sticky prices and high concentration in the Australian retail sector.

Two results are, viewed prospectively, surprising. It appears that issuers have chosen thus far to adjust their pricing structures mainly through fixed fees—i.e., fees that are independent of transaction volume. That is so despite a reduction in the variable revenue they received from the merchant side that had translated into negative variable prices on the cardholder side. Related to this is the finding that the near halving of interchange fees has not thus far led to a substantial reduction in card use.[64] That is to be expected for now given that marginal prices for completing transactions with cards have not changed significantly as a result of the regulation.

Over time we would expect a reduction in total credit-card use on the regulated side as a result of the decrease in the number of cardholders. However, it is difficult to predict the magnitude of that reduction without knowing the responsiveness of cardholders to increased membership fees. As noted, offsetting this decrease will be the increased use of cards from the unregulated proprietary systems. It appears that especially for high-spenders on rewards cards, issuers have made a serious, and predictable, effort to move them from the regulated associations to the unregulated proprietary systems. The RBA's hope that regulating the associations' interchange fees would exert substantial downward pressure on American Express and Diners Club's respective merchant discount rates is particularly unlikely to be realized as high-spend rewards cardholders move to the unregulated systems.

VII. CONCLUSIONS

Did the RBA regulatory intervention achieve its goals? One answer is that it is too soon to tell. It takes time for markets to adjust and it may be that banks and merchants will make further adjustment over time. But to the extent that the first couple of years provide information, the evidence indicates that the intervention has not achieved its goals for a reason that was not apparent to either the RBA or the card associations at the time the intervention was being debated.

[64] As we note above, growth rates of card output are lower post-regulation than pre-regulation. Although our prior is that this is a continuation of pre-existing declines in the growth rate, we do not rule out the possibility that growth rates might have levelled off absent regulation, and that the regulation has therefore lowered card growth.

The purpose of the intervention was to raise the cost of transacting with credit cards, to bring social costs and benefits in alignment, and thereby to reduce the use of what was thought to be an inefficient instrument. The RBA saw the interchange fee as the source of the problem because banks subsidized card use to get this revenue source and merchants could not make cardholders bear the differential costs of card use. If the intervention had resulted in banks raising usage fees in lockstep with the reduction in interchange fees or if merchants had raised surcharges to account for the interchange fees, the RBA would have achieved its objectives.

Neither expectation was fulfilled. For the most part, it seems that banks raised fixed fees and left the per-transaction incentives alone. In addition, the banks started to switch volume from the regulated to the unregulated systems. By and large, merchants have not chosen to impose surcharges. Thus far the incentives that cardholders face to use cards at the point of sale do not seem to have changed radically. If these patterns persist, the effects of the RBA regulation, if any, will likely take place through a reduction in total cards on the regulated systems.

This result does not just provide further support for the law of unintended consequences. It also raises interesting research questions. For the study of two-sided markets, it emphasizes that further work is needed to understand the role of two-part tariffs in guiding membership and usage decisions. This is an important topic for many two-sided industries since these two-part tariffs are quite common. For the study of regulatory interventions, it emphasizes that we need to know more about how two-sided businesses set prices to design interventions in those industries that can accomplish specified goals.

VIII. APPENDIX A

In order to determine whether issuers have made an attempt at recovering from cardholders some of the income lost due to the RBA regulation, we posed the question: In the period associated with the RBA regulation, is there evidence of a jump in the level of the card fees that issuers charge cardholders?

An examination of the Visa Australia series for other service charges per card and finance charges per card revealed that a few data points dramatically break the series trends. In the case of other service charges per card, the outliers are located in the second and third quarters of 1999 and the second quarter of 2000. In the case of finance charges per card, the outlier is located in the first quarter of 2004. To our knowledge, these are data problems rather than reflections of true change in the economic time series.

We took two different approaches in order to preclude our estimation from being driven by a few "influential" observations. First, we replaced the outliers with observations obtained via linear interpolation (Newton method). The results we report below were obtained on the basis of the interpolated data. Second, we kept the outliers in the dataset but ran robust regressions a la Huber (1964). In this iterative approach, observations receive weights that are a function of the magnitude of the associated residuals. For any given specification of the relevant model, the results obtained via robust regression were not substantially different from the ones obtained on the basis of the interpolated data.

A. Fee Levels

In order to explore the question regarding the jump in fee levels in the regulatory period, we estimated a model of the following form:

$$Y(t) = \alpha + \beta\, t + \gamma\, D(t) + \delta\, X(t) + \varepsilon(t) \tag{A1}$$

In equation (A1), Y(t) is the level of the fees that issuers charge cardholders in each quarter, t is a trend variable that takes on the value 1 at the starting point and grows by 1 each quarter, D(t) is a binary variable associated with the regulated period, and X(t) is a vector of other variables that may have had an impact on the level of Y(t).

More specifically, we want to isolate the behaviour of issuers—therefore, we have to control for those elements of the behaviour of cardholders that could have had an impact on Y(t). For example, in a model where Y(t) is real other service charges per card (including annual fees and other usage fees) we may want to control for real purchase volume per card (in case there are elements of Y(t) that vary with the level of purchase activity). In a model where Y(t) is real finance charges per card we will certainly want to control for the level of real outstanding balances per card. The trend variable picks up the effect of unobservable factors that may have changed over time and may have exerted an influence upon Y(t). We also include binary variables for the second, third, and fourth quarters in order to control for potential seasonality in Y(t).

We estimated a number of different models to probe how robust our results were to various specifications. In some of these models we took into account the fact that real purchase volume per card and real outstanding balances per card may be endogenous—in other words, cardholders may choose the optimal level of these variables by taking into account the issuers' choice of fees and interest rates. In other words, we checked whether we obtained considerably different results with OLS vis-à-vis an instrumental-variable approach. For space reasons, we do not report the results of all the estimated models here. The results not reported were not substantially different from the ones reported and they are available from the authors upon request.

In estimating models like (A1), researchers usually take two potential issues into account—the presence of serial correlation and the presence of stochastic trends. We dealt with serial correlation by estimating the models with robust standard errors—more specifically, Newey-West standard errors with four lags. We also used feasible generalized least squares (Prais-Winsten) estimation but obtained results not substantially different from those reported here. We checked for the presence of stochastic trends by running Dickey-Fuller tests and co-integration tests on the relevant variables.

1. Real other service charges per card

We first estimated models where Y(t) is real other service charges per card per quarter. Other service charges include annual fees and other service fees, like over-limit and late-payment fees. A regression of Y(t) on a trend variable, binary quarterly variables, and a binary variable for the regulated period that takes on the value 1 starting in the first quarter of 2003 produced a coefficient on the regulation binary variable of 3.99 (p-value = 0.000). Augmented Dickey-Fuller tests on the residuals rejected the null of a unit root with up to one lag. It is unclear whether the failure to reject the null with more than one lag was due to the lack of power of the test (due to the small sample size) or not.

We added the real sale volume per card on the right hand side and the coefficient on the binary variable dropped to 3.50 (p-value = 0.000). Augmented Dickey-Fuller tests on the residuals rejected the null (of no co-integration) with up to two lags. We then estimated a dynamic OLS (DOLS) model including the first difference of real sale volume per card on the right hand side, as well as one lag and one lead of the first difference. The coefficient on the binary variable dropped to 3.08 (p-value = 0.000).[65]

We followed a similar procedure starting with a regression of Y(t) (other service charges per card) on the trend variable, quarterly binary variables, and a binary variable that takes on the value one for the first time in the first quarter of 2004. The coefficient on the binary variable was 3.72 (p-value = 0.001). We added real sale volume per card, the first difference of real sale volume per card, and a lag and a lead of the first difference, and obtained a coefficient on the binary variable of 4.11 (p-value = 0.001).[66]

2. Consistency with RBA data

Our analysis based on quarterly Visa data is also broadly consistent with two separate sources of data on fees reported on an annual basis by the RBA.[67] The first source is total fee income (not including finance charges) received by banks on all credit cards. This series tracks closely to the Visa data. For example, average fee income per card in 2001 was AU$31.19 from the RBA data and AU$29.67 from the Visa data, and was AU$52.03 in 2003 from the RBA data and AU$52.45 from the Visa data. The increase from 2001 to 2003 was AU$20.84 from the RBA data and AU$22.78 from the Visa data.

The RBA also reports survey data on average fees for cards with interest-free periods and rewards programs issued by major banks. From 2002 to 2004, fees increased significantly: from $61 to $85 for annual fees on standard cards; from $98 to $128 for annual fees on gold cards; from $21 to $29 for late payment fees; and from $13 to $29 for over-limit fees. Thus, taking annual fees alone, for

[65] Incorporating the real 30-day interest rate on the right hand side generated an estimated coefficient of 3.18 (p-value = 0.002) for the reform binary variable.

[66] Incorporating the real 30-day interest rate on the right hand side generated an estimated coefficient of 3.90 (p-value = 0.001) for the reform binary variable.

[67] See RBA (2005a)

cards in this survey, fees increased by $24 a year for standard cards and $30 for gold cards. If we include one late fee and one over-limit fee penalty per year, total fees increased by $48 a year for standard cards and $54 a year for gold cards.

There are at least two reasons why these survey data on rewards cards indicate higher fee increases than the calculations we have reported based on the aggregate data (leaving aside the fact that the limited annual data series do no allow for full controls). First, cardholder spending is higher on rewards cards, so the loss in interchange fee income is greater. Second, replacing existing cardholders with new ones is costly and banks may be reluctant to raise fees as much to existing cardholders as to new cardholders.

3. Real finance charges per card

We followed the same approach with Y(t) defined as real finance charges per card. So, for example, a regression of real finance charges per card on a trend variable, a binary variable for the regulated period starting in 2003, and quarterly binary variables, produced a coefficient for the regulated period of 7.72 (p-value = 0.002). Dickey-Fuller tests did not reject the null of a unit root. After adding real outstanding balances per card on the right hand side, the coefficient for the regulated period dropped to 2.16 (p-value = 0.167). Augmented Dickey-Fuller tests rejected the null of no co-integration with two, three, and four lags. We estimated a DOLS model adding the first difference of real outstanding balances per card, and a lag and a lead of the first difference, and obtained a coefficient on the regulated period of 1.88 (p-value = 0.242).[68]

We followed the same approach starting with a regression of real finance charges per card on a trend variable, a binary variable for the regulated period starting in 2004, and quarterly binary variables. We obtained a coefficient on the regulation variable of 6.65 (p-value = 0.05), which dropped to 1.89 (p-value = 0.047) after controlling for changes in real outstanding balances per card. After adding the first difference of real outstanding balances per card, and a lag and a lead of the first difference, we obtained a coefficient on the regulated period of 1.14 (p-value = 0.316).[69]

4. Summary on fee levels

In summary, the evidence suggests that there was a jump in real other service charges per card associated with the RBA regulation. Depending on when we believe the structural break happened, the magnitude of the jump may have been roughly between AU$3 and AU$4 per quarter. The evidence of a jump in real finance charges per card is not as strong—the coefficients are much more sensitive to model specification and are, for the most part, not significant at standard confidence levels after controlling for changes in real outstanding balances per card.

68. Incorporating a measure of the opportunity cost of funds on the right hand side generated an estimated coefficient of 1.78 (p-value = 0.287) for the reform binary variable.

69. Incorporating a measure of the opportunity cost of funds on the right hand side generated an estimated coefficient of 1.25 (0.257) for the reform binary variable.

Part Five

SOFTWARE PLATFORMS

CHAPTER

14

Invisible Engines

David S. Evans

ABSTRACT

The use of software platforms to drive innovation and transform industries has exploded in the four years since the publication of the English-language edition of Invisible Engines in 2006. Around the globe, invisible engines are ushering in a new era of software-based technological change. The Apple iPhone has shaken the mobile phone industry worldwide in part by creating a massive applications business built on the phone's operating system. Firefox has revolutionized the browser industry by encouraging web developers to write add-ons and in doing so toppled Microsoft's Internet Explorer from dominance in many countries. Facebook has created a powerful social networking platform by opening itself up to developers. Amazon has released a cloud-computing platform that enables entrepreneurs to access its vast software, hardware and global communication systems over the Internet. A less well known company, IP Commerce, is starting to transform the payments business in the United States by helping developers build applications that work with the diverse hardware and software than handle the various types of payments.

In this preface to the 2010 Chinese edition of Invisible Engines we will survey what has happened since we finished the book and explain why software platform-based business models continue to create enormous social value, while often producing great profits for the companies and entrepreneurs behind them.

Invisible engines are based on software code. Software programs are written in various languages. Working by themselves or with other programs they tell hardware – from the pixels on your iPhone to the chip on your desktop to the communication devices providing your internet connection – what to do. They are the brains behind everything from the card reader in which you swipe your debit card to your social network page to the complex trading decisions at hedge funds to your favorite word processing package.

Software programs can become platforms that support other software programs when code that is written to perform a particular task could be made available to other software programs so the writer does not have to write that code again. That is what Windows does. It contains many little programs that help developers write their own programs. Microsoft provides a link to that code—called an Application Programming Interface (API)—that allows developers to link into the Windows code. Some businesses make their service available by providing an API. For example, YouTube provided an easy-to-use API for people to include access to videos on their social network pages.

By making APIs available, the owners of software programs are letting others use their intellectual property, the results of hard work writing code. They can benefit from doing this if others decide to write valuable applications that work with their software platform. More people are likely to use a platform if there are more attractive applications that work with it. For example, many users decided to switch from Internet Explorer to Firefox because there were add-ons for Firefox that did things that were not available with Internet Explorer such as support for XHTML – a popular programming language for webpage developers. Software platform providers do not have to make their code available for free though. Developers that sell their applications through the iPhone app store give a portion of their revenue (about 30 percent as of the end of 2009) to Apple.

Invisible engines are "catalysts" that use a multi-sided platform business model.[1] They ultimately create value by making it easier (usually by lowering the search or transactions costs) to get the members of different groups who value interacting with each other together. In the physical world, a shopping mall is a two-sided platform that helps bring shoppers and retail stores together. Software platforms generate value by reducing the cost for developers of writing applications for consumers and sometimes by providing a common place for users and developers to meet. For instance, the Chinese social network Renren is not only a virtual place for friends to get together, but also a place for friends to find applications that can help them and for application developers to find users. In this way, Renren creates value for its community of users and developers -and ultimately for itself.

The iPhone provides a potent illustration of this business model. By the end of 2009 there were more than 100,000 applications available for the iPhone. These applications programs have been downloaded more than 3 billion times since Apple opened its iPhone store in July 2008. Apple's success has shaken up the mobile phone industry in many countries around the world. It has unleashed tens of thousands of entrepreneurs and created a multi-billion dollar application industry for mobile devices. Here is how an invisible engine helped make this happen.

Inside the iPhone there is a computer chip on which is stored the software code that does all the magic that has been behind the success of this smart phone. That software code is usually called the operating system for the iPhone. Apple could have decided to keep that code entirely to itself and focus on making the iPhone a self-contained device that only did things that Apple wanted it to do. Instead, in a smart move, it decided to make that operating system a software platform

[1] David S. Evans and Richard Schmalensee, *Catalyst Code: The Strategies behind the World's Most Dynamic Companies* (Boston: Harvard Business School Press, 2007). Multi-sided platforms are sometimes referred to as "two-sided markets" even though they are usually businesses rather than markets.

that could support third-party applications. That meant it had to create APIs which would enable outsiders to link into the iPhone's operating system. These APIs give developers access to blocks of code that manipulate various aspects of the iPhone. Apple then provided a "software developer kit" (SDK) to help programmers develop applications using the iPhone's APIs and put together a web site to provide other resources for developers.

An important set of business dynamics starts when such an operating system is opened up and thereby becomes a software platform that can support developers. Apple initially provided more than enough applications and features on the iPhone to persuade many people to get one. By opening up its operating system it then started a virtuous circle. Entrepreneurs and hobbyists saw the opportunity to write interesting applications that could use the features of the iPhone, including its large touch screen and connectivity. The iPhone users soon found that they could download applications that made their phones even more useful than they were when they first bought them. More people started buying iPhones because of the increasingly valuable applications. As the number of iPhone users increased, developers had even more incentives to write applications.

Apple has not taken this process for granted. Like most other successful platforms it has stoked the fires. It developed a marketplace, an online shopping mall, to help developers make their applications available to iPhone users and earn revenues from them. Apple is responsible for collecting money from iPhone users and then gives the developers about 70% share of the revenue. Users have benefitted from having effectively a single location to search for applications. Developers have benefited because Apple has reduced the transactions costs of selling and has assembled an audience of users.

Apple has obviously made enormous profits and seen the market value of the firm increase because of the success of the iPhone and its application store. iPhone users have done very well also as a result of the tens of thousands of applications they can choose from, some of which enable them to do things that were not possible before. New markets have opened for entrepreneurs that use the iPhone, and some of them are earning significant profits or obtaining fame. By bringing developers and users together the iPhone has served as a catalyst which has created value—and profit—out of thin air. Others have followed suit. The most significant follower is Google, which has introduced the Android software platform for mobile phones, has stimulated hardware makers to introduce phones using this platform, and has worked hard at persuading developers to write applications for its store. China Mobile set up its own application store in the summer of 2009.

The invisible engine model followed by Apple's iPhone has helped power the information technology industry for about three decades. Apple itself was one of the pioneers in encouraging developers to write applications for its personal computer operating system. It invented the "software evangelist". Microsoft, though, was the maestro catalyst. Its Windows software platform attracted thousands of software developers and hundreds of millions of users. As a result it has been the durable center of a vast computer based ecosystem since at least the launch of Windows 3.0 in 1990. *Invisible Engines* tells the story of how software platforms have transformed industries including computers, video games, and handheld devices and then accelerated innovation. It then examines the forces behind the various business models that have been adopted in these industries. Many software platforms have decided not to charge developers while others have. Some have decided to vertically integrate into hardware, while others have remained pure software vendors.

The development of the internet and the spread of high-speed broadband throughout the world have generated a new wave of software platforms. Several of the major web businesses have turned themselves into platforms. The heart and soul of any web business does not reside in a server farm somewhere or in its buildings—it lies in the thousands and thousands of lines of software code that enable people to see and interact with web pages. Once written these software programs can be opened up to others by exposing APIs that enable developers to use portions of that code to interact with the web properties.

Facebook has allowed developers to write applications that work inside its social network. As a result developers can write games - such as Farmville – which Facebook users can play, or online shopping and advertising programs - such as those developed by FreeCause for raising money for charities. By the end of 2009 there were more than 500,000 applications running on Facebook. This five-year old site has also developed a program called Facebook Connect. Developers can write applications that through APIs made available by Facebook enable users of those applications to pull their Facebook identity, friends and privacy to other websites. As of the end of 2009 there were more than 6,000 developers working on Facebook Connect applications. These applications make Facebook more valuable to its users, and the greater number of users makes Facebook more valuable to developers. Facebook makes money in large part by selling access to the eyeballs drawn to Facebook and these applications to advertisers.

Amazon is one of several companies that are competing to create a general purpose software platform that can be accessed over the internet. Such platforms are said to reside "in the cloud" because they are off in the distance as opposed to on the user's desktop or other client computer. The idea is to develop a set of services that application developers can use. These include software services that allow developers to use the code rather than writing their own. But they also involve access to vast server farms and communication networks around the globe. Amazon's Web Services plays to this company's strength in selling products over the internet. It provides special services to developers who are writing programs to help merchants operate virtual stores and receive payment.

Invisible engines are beginning to disrupt the payments business both online and offline. This industry is at the heart of commerce and makes it possible for businesses and people to exchange value around the world. By definition it involves essentially all the money in the world. If invisible engines start driving innovation in this industry they could have an enormous impact on consumers and businesses who transact with each other.

An intricate set of systems and an enormous group of businesses are involved in payments. They range from the companies that print paper checks, to ones that make the point-of-sale terminals where you swipe your debit card, to banks that operate deposit accounts, to clearing and settlement networks. At the beginning of 2010 this industry involves many different computer systems, each with a software program, that have been made to interoperate with each other. Consider what happens when you swipe your credit card. The terminal sends your card information along with details of the transaction to a switch. That switch, which may be operated by a variety of businesses, has a software program that decides what to do with the transaction. It will send it on to a merchant processor which keeps track of the details of the transaction. This processor will also act as an intermediary with a clearing and settlement system, which will contact the bank processor that acts on behalf of the bank that issued the card (or possibly the bank directly). Many software programs are involved in the various steps of this process.

What we have just described is one "rail" in the payments system—the one for credit cards. While this may vary by country, in the United States there are other rails for debit cards, paper checks, electronic funds transfer, and other "tender" types. Now suppose that you are an entrepreneur who has come up with a great idea that involves incorporating payments into your application. This could be something as simple as a software program that helps small businesses accept payment in multiple ways and integrate these payments into basic accounting software. To succeed you would have to integrate your application into the many software programs that are used for the many tender types taken by businesses. That is such a daunting task—one involving substantial money and time not to mention the cooperation of the other businesses that handle payments—that you would give up.

This is a perfect problem for an invisible engine to solve. Massive transactions costs make it hard for payments entrepreneurs and payments users to get together. A software platform can lower those costs by investing in linking to the multitude of software programs that handle various elements of payments. By exposing APIs, this software platform then makes it possible for entrepreneurs to quickly integrate into most relevant aspects of the payments business. Such a software platform is analogous to Windows which, among other things, enables software applications to work with the multitude of device drivers controlling various hardware peripherals such as computer hardware, printers, and cameras.

In the United States several companies are working to create software platforms for payments. The most ambitious effort has been that of Denver-based IP Commerce. They have developed a Windows-like software platform enabling developers to link into software programs such as merchant processing platforms and point-of-sale equipment that account for the bulk of transactions in the United States. Developers have already used IP Commerce's invisible engine to create new products. PaySimple created the payments software for small businesses mentioned above. A small firm can install their web-based application and quickly accept payments offline and online with multiple tender types. Of course, IP Commerce is a new player and it remains to be seen whether it will succeed in generating the kind of virtuous circle that has made the fortunes of other software platforms and their creators.

There is nonetheless a great deal of evidence that the software platform model has caught on in payments. Amazon and PayPal are competing to offer payments platforms to merchants for online transactions. They have made software services available to developers through APIs. Application developers for merchants can use these APIs to make their applications capable of accepting payments using the PayPal or Amazon payments system. It remains to be seen whether the card networks will consider opening up their powerful software programs to outsiders and trying to build an ecosystem of applications around themselves.

Invisible engines are the unsung heroes of the economic progress that we have seen in the last three decades. It is true that these engines could not accomplish much without the massive improvements in microprocessor speeds, computer storage, and communication capabilities. But all those accomplishments would not have amounted to much without the development of software languages; the creation of software platforms that eliminate the duplication of effort in writing code; and the catalyst business model that enables the virtuous circle between users and applications.

Based on what has happened in the last four years it appears that we are entering a new age in the development of invisible engines. We suspect that historians will look back and define 1980-2004 as the first age of invisible engines. That is the period when software platforms helped create vast ecosystems around "thick clients" in which most of the work was done on a local computing device. The second age began after 2004 with the spread of Web 2.0, the rapid growth of smart mobile phones, and the global increase in broadband penetration. How long this second age will last we do not know, and it is even harder to guess what will come after it. With confidence, though, we can say that invisible engines will create new industries, destroy old ones, and drive innovation in China and elsewhere in the world.

David S. Evans
Andrei Hagiu
Richard Schmalensee
Boston, 23 January 2010

CHAPTER

15 Invinsible Engines: Both Sides Now

ABSTRACT

Two features of the technology we described in the last chapter shape the economics of software platforms. Software platforms are a written product of the mind. They are in effect documents, usually written in a high-level computer language. The code involved is malleable. It can be moved, altered, added to, and subtracted from with great ease. It is created almost entirely by people—"almost" because, like composers and writers, most programmers use computers for help. Software platforms are inherently multisided. They usually serve distinct groups of customers, who benefit from having each other on the same platform. Application Programming Interfaces (APIs) forge the crucial relationship between application developers and end users.

I'm basically a very lazy person who likes to take credit
for things other people actually do.

—Linus Torvalds[1]

INSIDE THIS CHAPTER

- Software platforms as information-goods and multisided platforms
- Economic and strategic characteristics of multisided platforms
- Economic aspects of open-source software

Two features of the technology we described in the last chapter shape the economics of software platforms.

Software platforms are a written product of the mind. They are in effect documents, usually written in a high-level computer language. The code involved is malleable. It can be moved, altered, added to, and subtracted from with great ease. It is created almost entirely by people— "almost" because, like composers and writers, most programmers use computers for help.

---

[1] Eric Raymond, *The Cathedral and the Bazaar* (Sebastapol, Calif.: O'Reily Press, 1999).

Software platforms are inherently multisided. They usually serve distinct groups of customers, who benefit from having each other on the same platform. Application Programming Interfaces (APIs) forge the crucial relationship between application developers and end users. The developer can benefit from using APIs when she can sell the resulting software to users who have those APIs on their computing devices.

Although we examine both of these features in this chapter, the multisided nature of software platforms is a main focus of the remainder of the book and the economic aspect of these invisible engines from which we will glean many insights. We conclude this chapter with a discussion of another remarkable aspect of the software business: people working collaboratively over the Internet, often without pay, produce software, including software platforms, that compete with software produced by for-profit firms.

I. INFORMATION GOODS

Software is one of many information goods in modern economies. Books, songs, screenplays, patents, and secret formulas are others. Of course, there are differences among these products. Most books are a final product read for enjoyment or knowledge. Musical scores instruct musicians on what to do with their instruments. And software is ultimately a series of instructions that directly or indirectly makes computer hardware work. But these differences pale next to the similarities.

Like all information goods, software has four major economic characteristics. It is a creation of the human brain; it is made of pliable symbols; the consumption of these symbols by one person does not exclude consumption by another; and it is almost costless to reproduce an exact replica of these symbols. A musical score has all these features. A composer can use musical notes to make scores of infinite variety and length. When an orchestra plays a particular score, it does not reduce the value of the score to anyone else. And it is cheap to reproduce the score, as well as any orchestra's rendition of it.

These software features have consequences shared by other information goods. Without intellectual property protection there is no obvious way to make money. (The free software movement discussed at the end of this chapter has found some unusual ways to motivate its participants.) There are extreme scale economies: fixed costs are high, marginal costs quite low. The addition of features is relatively easy and an important source of dynamic competition, incremental innovation, and product differentiation.

A. Software Characteristics

Produced by an Educated Workforce Software is designed and written by a diverse set of individuals, but typically they are college graduates who often have some training, and perhaps even a degree, in computer science. Software programmers and related professionals who worked in the U.S. software industry had an average of 15.3 years of education as of 2000. That compares with 13.8 years for the workforce on average and 14.7 years for professional service industries (including law, medicine, accounting, and architecture).[2]

There were 1,194 degree programs in computer science in American colleges and universities in 2006.[3] These programs usually offer courses in the design of operating systems. All of the top ten programs as ranked by *U.S. News and World Report* did.[4] There are a number of textbooks on the design of computer operating systems and related topics.

Microsoft is notable for screening people for intelligence and problem-solving skills. A 1995 study reported that Microsoft recruited from the top fifty colleges and universities and hired less than 3 percent of the people it initially interviewed. Microsoft is famous for asking job candidates to solve problems on the spot, such as estimating the number of gas stations in the United States. As of 2004, over 95 percent of the architects, designers, and programmers working on Windows had a college degree and 40 percent had computer science degrees. Ten years later Google has developed a reputation as a company where only brainiacs need apply.[5] It advertises mainly in technical magazines and puts people through numerous interviews that test intellectual skills before hiring one out of the 200 candidates who send in a résumé.

Made of Malleable Code We have already seen that software programs, including platforms, are a series of instructions usually written in a language such as C++. Managing the creation of millions of lines of code that work together as planned is no mean feat. But one reason these programs have grown so large is that they have been designed to make it easy to add to them. In some respects, doing so is like adding a paragraph to a chapter of a book or another riff when playing jazz. The key difference is that since the ultimate product is digital, users do not experience the addition in the way that adding a chapter makes a book thicker.

Like the contents of a newspaper, the contours of a software program can be changed readily. Just as newspapers have suburban or regional editions that add coverage specific to a particular geographic area, modern software programs may have versions targeted to particular groups, such as Java for small devices. And just as newspapers can add sections to bring in more readers, so software programs can add features or functionality.

---

[2] http://www.bls.gov/oco/ocos110.htm.

[3] http://www.usnews.com/usnews/edu/college/tools/brief/cosearch_advanced_ brief.php

[4] "America's Best Graduate Schools 2005 Edition," *U.S. News and World Report*, December 31, 2004.

[5] Michael A. Cusumano and Richard W. Selby, Microsoft Secrets (London: HarperCollins, 1996), pp. 92–93; Internal Microsoft information; John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture* (New York: Portfolio Press 2005); http://www.cbsnews.com/stories/2004/12/30/60minutes/main664063.shtml.

Many operating systems added features in the mid-1990s that helped users communicate over networks such as the Internet. Apple's Macintosh included AppleTalk proprietary networking protocols in 1985 and added protocols for communicating over the Internet in 1995.[6]

Easy to Reproduce All information goods are easy to copy. But software programs, including platforms, are especially so because they are necessarily digital. The easiest way to see this is with the open-source operating system Linux. You can download this 5.7 million-line operating system over the Internet from numerous Web sites.[7] With a cable modem connection it takes a couple of minutes. Not surprisingly, piracy is a major problem for software firms that sell their products.

Software platforms are often installed on computer hardware before it is sold. The manufacturer does this itself when it makes both the software and the hardware. Apple zaps its software right onto its iPods and Apple computers. Or the software manufacturer may license the software to other manufacturers that install it, often from a single master CD. Distribution costs are higher when the software platform is sold directly to users. Thus, Microsoft incurs some costs in reproducing its Windows software on CDs and distributing it through retail channels. The same is true for Linux distributors such as Red Hat. But even in these cases the per-unit costs are relatively low, as with other information goods such as music CDs.

Inexhaustible Once created, there is an inexhaustible supply of software such as platforms. Unlike most goods and services, but like all information goods, consumption by one user does not reduce the amount available for others. Indeed, software platforms are better than inexhaustible because consumption by one user is likely to increase the value of the software to others.

Complementarities and Network Effects System components are generally complements: adding another component to a system or improving an existing component generally increases the value of the other components. Moreover, in many cases, systems have what economists call *indirect network effects* linked to the presence of components.[8] That is, an increase in the number of users of one component often makes that component more valuable as a complement to the other components. As Sony's Internet–based game center for the PlayStation 2 draws more users, for instance, more PlayStation 2 owners will want to buy games supported by the Internet service. As its games become more popular, more consumers will prefer PlayStation 2 consoles to competitors'

[6] Jim Carlton, Apple: The Inside Story of Intrigue, Egomania, and Business Blunders (New York: HarperCollins, 1997), p. 59; "Apple goes to the core; Apple introduces Power Macintosh 9500 that uses TCP/IP and PCI bus architecture; Product Announcement; Brief Article," LAN Magazine, October 1, 1995.

[7] http://zdnet.com.com/2100-1104-864256.html.

[8] Michael Katz and Carl Shapiro, "Systems Competition and Network Effects," *Journal of Economic Perspectives* 8 (Spring 1994): 99. For a general discussion of network effects and their business implications, see Carl Shapiro and Hal Varian, Information Rules (Boston, Mass.: Harvard Business School Press, 1998).

models. Likewise, an increase in the variety of components (printers for PCs, for example) often increases the value of other components to end users. In recent years economists have tended to apply the multisided platform framework to these situations, as we discuss in the next section.

There also may be direct network effects. These arise when an increase in the number of users of an application or platform directly makes that application or platform more valuable to each user. Its value increases because users can share information and work together more efficiently. When WordStar was the leading PC word-processing program, many people bought it at least in part because they could share documents with friends and co-workers.

B. Economic Consequences

These technological features shape the economics of software platforms just as they shape the economics of most software products.

Intellectual Property Protection If people could get the source code for any software product, they could reproduce it for next to nothing. The price would fall to almost zero, and the original writer would derive no financial benefit.

Software companies rely on all three major forms of intellectual property protection to guard their investments against this fate.[9] First, they keep the source code secret as much as possible. Before they distribute the software they turn it into the 1s and 0s of machine code. In principle, an able, dedicated, and patient programmer could translate machine code back into a high-level language. But this sort of decompiling is forbidden by almost all commercial software licenses and all but impossible in practice for multi-million-line software platforms. In addition, "trade secret" law protects software developers from the unscrupulous employee or agent who might try to release the source code without authorization.

Software companies also copyright their code. As with a book, you cannot reproduce copies of software programs without violating copyright law. Of course, as with other information goods, piracy is nonetheless rampant, especially in countries with weak intellectual property laws. In India, an estimated 80 to 85 percent of the copies of Macro media Flash and Dreamweaver in use are not legal. Even in the United States, almost 30 percent of Macromedia software is pirated.[10]

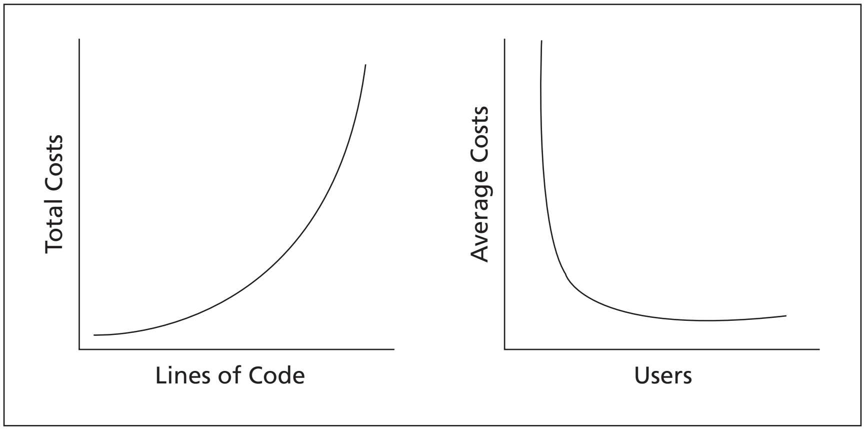
Finally, software companies get patents on algorithms and other features. The United States had granted about 127,000 software patents through 2004. (It takes about 3.5 years on average for the grant of a patent application.[11]) Apple's iTunes software, for example, allows users to import

[9] Suzanne Scotchmer, Innovation and Incentives (Cambridge, Mass.: MIT Press, 2004), chap. 3.

[10] Venkatesh Ganesh, "The continuing story of software piracy," *The Financial Express* (August 8, 2005), http://www.macromedia.com/devnet/logged_in/ swozniak_piracy.html.

[11] It is difficult to define "software patents," but in 2004, 127,098 patents were granted under the G06F classification, which covers electronic data processing. Data set available from www.uspto.gov.

Figure 1 Diseconomies of scale.



an unlimited number of audio tracks and encode them into the popular MP3 format, as well as listen to MP3s, audio CDs, or hundreds of Internet radio stations. A patented system for accessing digital media across networks was important for the success of iTunes.

Economies and Diseconomies of Scale Although no one has ever quantified it, it is generally understood that there are diminishing returns to scale in writing software platforms. That is, doubling the size of a platform by adding more features more than doubles the cost (holding the quality of the code constant—one can always write inefficient code). Increases in size create more interdependencies (with N objects, there are N2 possible pairwise interactions, for instance), thus raising the likelihood of bugs, and thereby raising development, debugging, and test costs more than proportionately.[12] (Object-oriented programming and the use of modules are designed to temper these diseconomies.) But once created, cheap reproduction means that additional copies cost little. The production technology is therefore as shown in Figure 1.

These economic features suggest some caution in characterizing the marginal cost of producing software. It is true that once the costs of creating software have been sunk, the marginal costs of reproducing and distributing it are very low. That is an ex post perspective on cost. But it is also true that the likely adoption of a software program is not independent of the costs that are incurred in creating it or revising it. Software designers add features in part to bring in more consumers; in the case of software platforms, those consumers include both users and developers of applications that run on top of the platforms. Ex ante, the marginal cost of acquiring additional customers by improving the platform, is likely to be much higher than the marginal cost of reproducing and distributing the software.

[12] "Linux: Fewer Bugs Than Rivals," *Wired News*, December 14, 2004.

This distinction is also relevant for other information goods. Movies are a good example. The cost of making a completed movie available to an additional viewer is close to zero. However, the number of viewers a movie garners is partly dependent on the investment in the actors, special effects, and other features that make a movie popular. Movies made with low budgets are often aimed for a narrow audience, while movies with blockbuster potential typically have extravagant budgets. The marginal cost of garnering a viewer, viewed ex ante, is positive.

Pricing and the Recovery of Investments As with all information goods, software poses some challenges related to recovering investments and earning profits. Pricing at ex post marginal cost or anything close to it would lead to bankruptcy. Software pricing thus depends primarily on demand (particularly the responsiveness of demand to changes in prices) rather than on cost and has as its main goal at least recovering fixed and sunk development costs. The pricing of software platforms is considerably more complex because of their multisided nature, as we discuss later.

Why have software prices not declined at the same pace as hardware prices? Basically because software development costs have not declined as rapidly as hardware costs, for two related reasons. The first is that educated labor, which is not becoming cheaper, accounts for most of the cost of producing software. In 2001, U.S. software firms paid about 33 percent of revenues to their employees, while semiconductor companies paid less than half that percentage.[13] The second reason is that software products are becoming more complex: with advances in hardware, software programs typically grow over time through the accretion of features. A typical PC game program in 1994 was 20 megabytes; a typical PC game program in 2004 was 2,200 megabytes.[14]

Bundling Features Most goods are bundles of features, many of which could be provided separately but are not. Cars come with spark plugs and tires even though you could buy your own. Moreover, many goods are improved over time through the addition of features. Few cars come these days without air conditioners and rear window defrosters.[15] Many cereals add fruit and flavors over time, leading to many variations.

The same is true for computer systems. Microprocessors, memory, and other components are typically combined to create a hardware platform such as a Nokia mobile phone handset or an Xbox game console. With time, many peripherals come to be integrated into the hardware platform. Consider the case of the math coprocessor, which facilitates number crunching. Before the release of Intel's 486 chip, Intel's microprocessors did not include a math coprocessor. Customers

[13] 2002 Economic Census, Industry Series Reports, available at http://www. census.gov.

[14] "Doom II Game Has 500,000 Pre-Orders," *Newsbytes News Network*, October 13, 1994; http://www.amazon.com/exec/obidos/tg/stores/detail/videogames/ B00006C2HA/tech-data/103-8949918-3544622.

[15] David Evans and Michael Salinger, "Why do Firms Bundle and Tie? Evidence from Competitive Markets and Implications for Tying Law," *Yale Journal on Regulation* (January 2005): 37–89.

who wanted one purchased it separately from one of several vendors at substantial cost. Today, one cannot buy an Intel x86 processor without a built-in math coprocessor.

Creating products through feature addition is particularly easy with information goods. That is the beauty of the pliability of music and language. Pop music was mainly distributed as singles on seven-inch records until the early 1960s. The success of the Beatles' *Sergeant Pepper's Lonely Hearts Club Band* album made clear the value of compiling songs and helped make the market for long-playing albums. Bundling multiple songs into albums became standard practice, and the distribution of songs as singles became less common. Newspapers have added various features such as style and living sections over time.

Similar forces apply to software in general and platforms in particular. Where exactly the tasks performed by software are accomplished is a matter of business and design decisions. Many tasks that used to be performed by stand-alone applications have become integrated into other applications (such as spell checkers, which originally were sold separately from word processing programs) or into the software platform itself. Early operating systems, for example, did not include communications functionality.

The malleability of code reinforces several economic forces that encourage the inclusion and accretion of features in products.

Bundling and integration. Combining features in a single product reduces transaction costs for consumers. Rather than having to buy two products, they can buy just one. Moreover, the manufacturer can create additional value by creating connections between the features. An example is making the features of a spreadsheet program available to a word-processing program.

Economies of scope. When there are fixed costs in offering separate products, firms may find it profitable to bundle those products if demand for the separate components is not particularly strong. Several major automobile makers, for example, have decreased the number of different cars people can purchase.[16] They have done this by developing bundles of options that "most people" want, even though some people would not value some of those options separately. For software there are cost savings from combining several features into a single package, as well as savings in distribution and product support. (There may be diseconomies, of course, if making the program larger and thus more complex results in disproportionately large increases in the costs of writing, debugging, testing, and supporting the package.)

Demand aggregation. When there are fixed costs of producing and distributing products but low marginal costs of adding components, it may be possible to lower average costs and reduce variation in what people are willing to pay by combining components that appeal to different groups of customers.[17] Hardware and software typically include many features that most con-

[16] Ibid.

[17] Yannis Bakos and Eric Brynjolfsson, "Bundling and Competition on the Internet," Marketing Science, 1 (Winter 2000): 63–82.

sumers never use. However, by including these features vendors expand the number of consumers who find the product valuable at the offered price. This is why many word processors include equation editors, newspapers have horoscopes, and cable companies include channels that most of us never watch.

II. MULTISIDED PLATFORMS

That shopping malls and software platforms have much in common is one of the important insights of the economics of multisided markets. The mall is available to stores and shoppers. Once there, the merchants and consumers interact directly on the platform. The merchant rents space. The shoppers often get amenities such as free parking, in addition to getting into the mall for free.

Likewise, the software platform is available to developers and users. The developer licenses its software to the user, who then runs the application on the platform. Both user and developer rely on the services provided by the platform. For many software platforms the user pays to license the platform, while the developers get to use the platform services for free and may even get some subsidized software tools to help them do so.

Both platforms help reduce duplication and thereby lower the cost of providing services. Shopping malls provide parking, restrooms, and many other common facilities. Stores benefit because they do not have to provide these facilities on their own. Shoppers benefit because retailers have lower costs. Software platforms make services available through APIs. Developers benefit from these because they avoid having to write some of their own code. Users benefit from a greater variety of and lower prices for applications.

The economics of multisided platforms provides a set of tools for understanding the past, present, and future of software platforms.

A. The Economics of Multisided Platforms

Multisided platforms cater to two or more distinct groups of customers. Members of at least one customer group need members of the other group for a variety of reasons. Platforms help these customers get together in a variety of ways and thereby create value that the customers could not readily obtain otherwise. The village market is one of the oldest examples of a two-sided platform. It is a place where buyers and sellers can get together and trade. So is eBay. Another old example is the village matchmaker, who helped men and women find marriage partners. Match.com provides a similar service using Internet technology; speed dating is another important innovation. The publisher of this book operates a platform, too. It is in the business of finding authors in search of an audience and audiences in search of content.

Governments run some two-sided platforms. Cash is an example. The government institutions behind the euro help ensure that sellers will take it for payment and buyers will use it for payment. Standards sometimes give rise to two-sided platforms. Fax machines facilitate communication between senders and receivers. Cooperatives of firms also operate two-sided platforms—Visa is the most significant example. For-profit businesses operate two-sided platforms in a wide variety of industries and in many economically significant ones. Highly visible examples include American Express (travelers checks and charge cards), Google (search engine–based portal), and News Corporation (advertising-supported media).

William Baxter presented one of the first formal analyses of a two-sided business in 1983.[18] He was a law professor who was self-taught in economics. He observed that payment cards provided a service only if both cardholders and merchants jointly agreed to use a card for a transaction. He derived some of the fundamental economic consequences of this joint demand. (Baxter went on to become a highly innovative antitrust chief in the United States.)

The notion, however, that diverse industries are based on two-sided platforms and are governed by the same basic economic principles is due to a pathbreaking paper by Jean Tirole and Jean-Charles Rochet that began circulating in 2001.[19] They showed that businesses such as computer operating systems, dating clubs, exchanges, shopping malls, and video game consoles were two-sided.

Economists now recognize that many industries, including the manufacture of software platforms, are guided by economic principles that differ in important ways from those that govern traditional industries. Many of these two-sided or multisided industries are subject to network effects, which were studied extensively by economists during the 1980s and 1990s.[20] Network effects are also central to the economics of multisided platforms, and more recent analysis provides additional insight into their business implications.

Internalizing Externalities Multisided platform businesses tend to arise in markets that have three characteristics:

1. There are two or more distinct groups of customers.
2. There is some benefit from connecting or coordinating members of the distinct groups.
3. An intermediary can make each group better off through coordinating their demands. For example, dating clubs provide a service to men and women, who benefit from meeting each other, and provide an efficient way for men and women to connect.

[18] William Baxter, "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *The Journal of Law and Economics* 26 (October 1983): 541–588.

[19] Jean Tirole and Jean-Charles Rochet, "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association* 1, no. 4 (2003): 990–1029.

[20] Shapiro and Varian, Information Rules. See the articles in Symposium on Network Externalities, *Journal of Economic Perspectives*, 8 (Spring 1994): 93–150.

As a practical matter, multisided platforms tend to arise when a stronger version of condition 2 applies: most platform businesses exhibit indirect network externalities. Consumers, for example, get more value from their credit cards when more merchants take them, and merchants get more value from accepting credit cards when more consumers use them. This has not been lost on the card systems. The current advertising slogans highlight merchant acceptance: "Visa. Everywhere you want to be." "MasterCard: No card is more accepted." American Express, MasterCard, and Visa persuade merchants to pay for taking their cards by emphasizing the millions of consumers that have these cards and want to use them to pay. The sales pitch for the merchants is similar: then the card systems tout the number of cardholders they have who could transact at the merchant if it accepted the card for payment.

Customer groups can sometimes get together without a platform. Men and women have found each other without matchmakers. Buyers and sellers figured out ways to transact before there was money. Merchants can advertise their wares without the media. Successful multisided platforms, however, generally reduce the transactions costs that members of different customer groups would incur in trying to reap the benefits of getting together.

The fact that a platform could exist does not mean that it necessarily will or that it will provide the only method for providing benefits to customers. Apple has thus far operated its iPod/iTunes platform as a single-sided business.. It buys music by paying publishers royalties and distributes this music to customers who want it. Similarly, many consumers have "store cards"—payment cards issued by stores such as Bloomingdale's. In fact, the payment card industry was based entirely on this single-sided model until Diners Club introduced a card in 1950 that put multiple merchants and consumers on the same platform.

Similarly, many businesses deal with multiple diverse groups without being platforms. There is a sense in which auto companies bring tire manufacturers and consumers together, but they do not do so in a way that makes Toyota, for example, a multisided platform business. In this case there is no direct interaction between the two sides. Toyota substitutes itself for consumers when dealing with tire producers, just as Apple does before sending music to consumers through iTunes. By contrast, two-sided platform businesses provide support for direct interaction between the two sides. Thus, game developers sell directly to PlayStation users, for instance, not through Sony.

Multisided businesses can generate profits for themselves and benefits for their customers if they can figure out ways to increase and then capture indirect network externalities. There are three major ways in which they do this.

First, they serve as matchmakers. Financial exchanges such as NASDAQ and on-line auction sites such as eBay match buyers and sellers. The Yahoo! Personals and 8MinuteDating match men and women.

Second, they build audiences. Advertising-supported media do mainly that: they use content to attract eyeballs and then sell access to those eyeballs to advertisers. Many platforms engage in less overt audience building. Auction houses such as Sotheby's try to build an audience of buyers for the art they sell on consignment. Nightclubs sometimes try to build an audience of women for men, and vice versa. We saw that payment card systems try to build an audience of cardholders for merchants and an audience of merchants for cardholders.

Third, they reduce costs by providing shared facilities for the customers on each side. That's the shopping mall case with which we began. But other platforms also do this to some degree. Buyers and sellers have shared auction institutions and auction sites from the Roman forum to eBay. Readers and advertisers share the pages of Vogue. Payment card systems provide a shared network for conducting transactions between merchants and consumers.

Software platforms provide value through matchmaking and building audiences, as well as through reducing duplication. Apple, for example, helped bring commercial artists and developers of design software together. It did this by including services in the Mac OS that developers could use to develop programs such as Adobe Photoshop for commercial artists. Sony PlayStation has developed an audience of console users that it can make available to game developers. The main economic value of software platforms, however, is in economizing on the amount of code that developers must write to serve the needs of consumers.

The Pricing Balancing Act In single-sided markets, price usually tracks costs and demand for the product pretty closely. Firms figure out what their marginal cost will be and then mark it up—a little if customers are price-sensitive because there is a lot of competition, more if there is little competition. Particularly in stable markets, this is not rocket science. That is why how-to books on starting your own small business can offer reliable advice, such as "charge X times cost in sector Y." For example, one guide advises that the markup is generally 40 percent of the retail price in hardware stores and that for jewelry it ranges between 400 and 800 percent.[21]

In multisided markets, pricing is more complicated because of indirect network effects between the distinct customer groups. If you charge women the same price as men to enter your singles club, you may not get enough women. If this happens, men will not come, and suddenly you will have an empty club. Many Internet publications discovered that viewers deserted in droves when they attempted to charge them, although some did make the successful transition to paid subscriptions plus advertising.[22]

---

[21] Stephen C. Harper, *The McGraw-Hill Guide to Starting Your Own Business*, 2nd ed. (New York: McGraw-Hill, 2003), pp. 100–101; Jan Kingaard, *Start Your Own Successful Retail Business* (Irvine, Calif.: Entrepreneur, 2003), pp. 152–153.

[22] Michael Liedtke, "Online Subscriptions Herald the End of Web Freedom," Associated Press newswires, March 18, 2002; Thomas E. Weber, "Web Users May Balk at New Fee Services That Deliver Little Value," Wall Street Journal, April 8, 2002; Timothy J. Mullaney, "Sites Worth Paying For? The Paid Web Is a Work in Progress, But Some Are Already Getting It Right," *Business Week*, May 14, 2001.

Table 1 Revenue in Selected Multisided Platforms

Industry	Multisided Platform	Sides	Side That Is "Charged Less"
Real estate	Residential property brokerage	• Buyer • Seller	Buyer
Real estate	Apartment brokerage	• Renter • Owner/landlord	Typically renter
Media	Newspapers and magazines	• Reader • Advertiser	Reader
Media	Network television	• Viewer • Advertiser	Viewer
Media	Portals and Web publications	• Web "surfer" • Advertiser	Web "surfer"
Finance	Proprietary terminals	• Trader/analyst • Content provider	Content provider
Shopping malls	Mall	• Merchant • Shopper	Shopper
Payment system	Travelers' checks	• Check holders • Merchant	Merchant
Payment system	Charge/debit card	• Cardholder • Merchant	Cardholder

Multisided platform economics shows that it may make sense for firms to charge very low prices to one or more groups or even to pay them to take the product. And that is what multisided businesses do. Magazines, newspapers, and television broadcasters typically earn the preponderance of their revenues from advertisers.[23] Charge card companies such as American Express earn the bulk of their revenue from merchants.

Businesses in multisided markets often subsidize one side of the market to get the other side on board—sometimes explicitly by charging low or negative prices. At other times subsidies are less apparent, such as when the platform makes significant investments in one side and does not charge for it. Table 1 shows some examples. We will see that all software platforms make services available to at least one side for free. Most make free services available to developers through the APIs.

The economics of pricing for multisided platform businesses has another key implication. In single-sided businesses, the principle that the one who causes the cost should pay the cost is good

[23] Lisa George and Joel Waldfogel, "Who Benefits Whom in Daily Newspaper Markets?" NBER Working Paper no. 7944 (October 2000), p. 9.

advice, for businesses as well as for policymakers. For example, a car buyer "causes" the cost of manufacturing the car, and thus pays the full cost. That principle usually does not make any sense in multisided markets, however. Often a product cannot exist unless several different customers participate simultaneously. They all "cause" costs and "cause" benefits. That is true even if it is possible to identify costs that increase as a result of an additional user on one side—for example, the cost of printing another copy of the Yellow Pages. Economists have shown that the best prices—either from the standpoint of the business maximizing profits or from the standpoint of policymakers maximizing social welfare—involve complex relationships between the price sensitivity of each side, interdependencies between these demands, and marginal costs.

Is There Anything New Here? Multisided platforms have a number of features that economists have examined before. Yet traditional learning does not deal with the role of intermediaries in internalizing network externalities. Most businesses have distinct consumer types: workers or retirees, households or corporate entities, men or women. But multisided platforms differ in that they must serve two or more distinct types of consumers to generate demand from any of them. Hair salons may cater to men, women, or both. Heterosexual dating clubs *must* cater to men and *women*. For hair salons the ratio of men to women does not matter much; for heterosexual dating enterprises it is absolutely critical.

Most businesses in single-sided and multisided markets engage in price discrimination (charging different prices that aren't proportional to the corresponding marginal costs) because it is possible to increase revenue by doing so and because, in the case of businesses with extensive scale economies, it may be the only way to cover fixed costs. A dating club may charge men a higher price just because they have more inelastic demand and because it is easy to identify that group of consumers. But businesses in multisided markets have an additional reason to price discriminate: by charging one group a lower price the business can charge another group a higher price; and unless prices are low enough to attract sufficient numbers of the former group, the business cannot obtain any sales at all.

Like firms in multisided markets, many firms in single-sided markets sell multiple products, and there is extensive economic literature explaining why they do so. The standard explanations for why firms produce multiple products probably apply to many of the platforms discussed here. But firms that make multiple products for several one-sided markets (for example, General Electric makes light bulbs and turbine engines) or several complementary products for the same set of consumers (for example, IBM sells computer hardware and computer services) do not secure profit opportunities from internalizing indirect network effects.

Finally, it is important to ask how the business implications of the recent work on multisided markets differ from those of the older economic literature on network effects. This is not as simple as it might seem, since popular discussions of network effects often missed important subtleties in the academic literature.

Take the case of single-sided markets with direct network effects. Because of those effects it follows that there is an advantage to size, all else equal. But it does not follow that this advantage, if present, is large, and it certainly does not follow that the firm with the biggest market share always wins in the end, let alone that the first entrant always wins. Nonetheless, in the frenetic days of the Internet bubble, lots of businesses were founded on the assumption that network effects were present and important in their markets and that the key to success was to get in fast, price low, and build share at any cost. Proponents of this simplistic view emphasized tipping—you build up critical mass, and then the whole market flocks to you. And they emphasized an extreme sort of lock-in— once you get most of the customers, nobody can enter against you, even with a better product.

As Brian Arthur, an author of several influential papers in network economics, put it, "You want to build up market share, you want to build up user base. If you do you can lock in that market."[24] This is a nice, simple theory—much simpler than the economic literature from which it claimed to be derived. But it is hard to find many businesses that succeeded by following it. Unfortunately, many dot-com entrepreneurs and investors thought that "build share fast" was the path to great riches. Only a few made it very far down that path before reality closed it off and supposedly locked-in buyers left en masse. It turns out that only rarely are direct network effects strong enough to prevent buyers from switching to a better product, as the massive defections of buyers from once dominant word-processing programs illustrates.

Those who believed that riches could be made quickly and easily by harnessing network effects tended not to distinguish sharply between direct and indirect network effects. In both cases the managerial prescription was to build share rapidly; indirect network effects, like direct network effects, would kick in automatically and both fuel and protect further growth. Work on two-sided markets makes it clear that this is dangerously simplistic in two important respects.

First, even though at least two distinct groups must be involved or there to be indirect network effects, the network enthusiasts assumed both that it is obvious that one should pay and the other should not, and that it is obvious which group should pay.

Second, they generally assumed that the group that did not pay could be ignored in setting business strategy because it would automatically fall into line and generate valuable network effects. In contrast, economic analyses of multisided platforms, show that successful multisided platform businesses must pay careful attention to all relevant groups, and typically must worry more about balance among them than about building share with one of them. The multi-sided approach is consistent with asymmetric treatment of customer groups, but getting it right requires great luck or careful analysis.

[24] Joel Kurtzman, "An Interview with W. Brian Arthur," *Strategy+Business* 11 (1998): 100.

The popular network economics literature also suggested that markets with direct or indirect network effects would tend to tip toward a single provider.[25] That does not happen much in practice, though. Sometimes congestion costs outweigh network effects—that is the case with nightclubs, trading pits, and shopping malls. Platforms also differentiate themselves, and thereby counter the network effects of their rivals, by trying to appeal to different consumer preferences. That is one of the reasons for the proliferation of magazines.

Consider some markets that seem to display important indirect network externalities: PC operating systems, real estate agencies, payment cards, auction houses, local and national newspapers, broadcast networks, parcel delivery services, banks, dating services, standards for encoding DVDs, financial information services, music publishers, and recorded music manufacturers. Of these many markets, the only ones in which a single large player accounts for the preponderance of sales are PC operating systems (i.e., Windows) and some local newspaper markets (such as the *Los Angeles Times*).

Most software platform categories are competitive as a result of providers differentiating themselves to appeal to different types of customers on either side of the market.

Business Models in Multisided Platform Markets Making a platform a success is a delicate process. Businesses have to get the pricing structure right; they must balance the demands of the various customer groups and nurture the several sides of the market. Getting the balance right seems to be more important than building shares. Platform markets do not tip quickly because as a practical matter, it takes time to get things right. And the first entrant often does not win in the end: many other firms may come in and successfully tweak the pricing structure, product design, or business model. eBay is a successful business-to-business (B2B) exchange now, for example, but many earlier B2Bs failed.[26] Most B2Bs tried a big-bang strategy: make substantial investments in a platform and hope both sides show up when the platform opens for trading. The first and third

[25.] Some of the network effects models allowed for differentiated tastes and the coexistence of multiple networks. See Jeffrey Church and Neil Gandal, "Network Effects, Software Provision and Standardization," *Journal of Industrial Economics* 40, no. 1 (March 1992): 85–104. S.J. Liebowitz and S.E. Margolis, "Network Externality: An Uncommon Tragedy" *Journal of Economic Perspectives* 8 (1994): 133–150. However, this literature was often taken to suggest overall that network effects naturally lead to a single firm dominating a category. See Brian Arthur, "Increasing Returns and the New World of Business," *Harvard Business Review* 74 (July–August 1996): 100–109.

[26] eBay was not begun as a B2B Web site, but as more and more businesses began to do business on it, it became one. In 2003, eBay officially launched a separate B2B site. "Prior to the B2B site, eBay listed more than 500,000 business items for sale every week on its consumer site, with business buyers representing more than $1 billion in annualized gross merchandise sales, officials said." Renee Boucher Ferguson, "eBay Launches B2B Site," *eWeek*, January 28, 2003 (http://www.eweek.com/print_article/0,3048,a=36363,00.asp); eBay Press Release, "eBay Launches eBay Business to Serve Its Growing Community of Business Buyers," January 28, 2002 (http://investor.ebay.com/news/20030128100772.cfm); Mark Berniker, "SAP, eBay Setup Industrial B2B Marketplace," *internet News.com*, June 16, 2003 (http://www.internetnews.com/xSP/article. php/2222371).

entrants into the payment card industry, Diners Club and Carte Blanche, barely exist today. The second entrant, American Express, had a 14 percent share of credit and debit card purchase volume in 2003.[27]

Getting All Sides on Board An important characteristic of multisided markets is that the demand on each side vanishes if there is no demand on the others, regardless of what the price is. Merchants will not accept a payment card if no customer carries it because no transactions will materialize. Computer users will not use an operating system for which the applications they need to run have not been written (except those rare users who plan to write their own applications). The businesses that participate in such industries have to figure out ways to get both sides on board.

One way to do this is to obtain a critical mass of users on one side of the market by giving them the service for free or even paying them to take it. Especially at the entry phase of firms in multisided markets, it is not uncommon to see precisely this strategy. Diners Club gave away its charge card to cardholders at first—there was no annual fee, and users got the benefit of the float.[28] Netscape gave away its browser to many users, particularly students, to get a critical mass on the end-user side of its business.[29] (Initially the other side was providers of Web sites, to whom Netscape sold its server software.)

Another way to solve the problem of getting the two sides on board simultaneously is to invest to lower the costs of consumers on one side of the market. As we saw earlier, for instance, Microsoft invests in the creation of software tools that make it easier for application developers to write application software for Microsoft operating systems and provides other assistance that makes developers' jobs easier. In some cases, firms may initially take over one side of the business in order to get the market going. Palm would never have succeeded in creating the vibrant Palm economy, with thousands of software applications and hardware add-on developers and millions of users, had it not provided the first applications itself (especially Graffiti, the handwriting recognition system).[30]

Providing low prices or transfers to one side of the market may help the platform solve the simultaneity problem by encouraging the benefited group's participation—which in turn, owing to network effects, encourages the nonbenefited group's participation. In addition, providing benefits to one side can discourage its use of competing multisided platforms. For example, when Palm provides free tools and support to PDA applications software developers, it encourages those

[27] Nilson Report 805 (February 2004).
[28] "Credit Cards for Diners," *New York Times*, March 30, 1950, p. 37; Diners Club display advertisement, *New York Times*, March 30, 1950, p. 42.
[29] David Plotnikoff, "Internet Born with Netscape," *Mercury News*, February 28, 2003 (http://www.tc.umn.edu/~jbshank/7_NetscapeIPO.html).
[30] Annabelle Gawer and Michael Cusumano, *Platform Leadership: How Intel, Microsoft and Cisco Drive Industry Innovation* (Boston: Harvard Business School Press, 2003).

developers to write programs that work on the Palm OS platform and automatically induces those developers to spend less time writing programs for other operating systems.[31]

Pricing Strategies and Balancing Interests Firms in mature multisided markets—those that have already gone through the entry phase, in which the focus is on getting the various sides on board—still have to devise and maintain an effective pricing structure. In most observed multisided markets, companies seem to settle on pricing structures that are heavily skewed toward one side of the market, as Table 3.1 shows. Google earns the preponderance of its revenue from advertisers, for instance, and real estate brokers usually earn most or all of their revenues from sellers.

Sometimes all competing platforms converge on the same pricing strategy. In principle, Microsoft, Apple, IBM, Palm, and other operating system companies could probably have charged higher fees to applications developers and lower fees to hardware makers or end users. Most discovered that it made sense to charge developers relatively modest fees for developer kits and, especially in the case of Microsoft, to give away a lot for free.

Getting the pricing balance right, however, requires considerable care. For example, in 2000, Yahoo!'s Internet auction site was second only to eBay in terms of the number of listings. Sellers found the site appealing because unlike eBay, Yahoo! did not charge sellers a fee for listing their products. In 2001, Yahoo! changed its pricing strategy and began charging a fee. Yahoo!'s listings dropped by 90 percent as sellers presumably moved to the larger venue, eBay.[32] The price change affected Yahoo!'s buyer-side market as well, since buyers were now left with little to bid on.

Two important factors influence multisided pricing structures. There may be certain customers on one side of the market—Rochet and Tirole refer to them as "marquee buyers"[33]—who are extremely valuable to customers on the other side of the market. The existence of marquee customers who create strong network effects tends to reduce the price to all customers on the same side of the market and increase it to customers on the other side. A similar phenomenon occurs when certain customers are extremely loyal (or captive) to the multisided platform firm, perhaps because of long-term contracts or sunk cost investments. The effect is then opposite: the presence of captive customers leads to an increase in the price charged to those on the same side and a decrease in the price charged to the other side.

For example, American Express has been able to charge a relatively high merchant discount as compared to other card brands, especially for its corporate card, because merchants have viewed the American Express business clientele as extremely attractive.[34] Corporate executives on expense

---

[31] Jean-Charles Rochet and Jean Tirole, "Platform Competition in Two-Sided Markets," working paper, December 13, 2002; http://www.palmsource.com/ developers/why_develop.html.

[32] Saul Hansell, "Red Face for the Internet's Blue Chip," *New York Times*, March 11, 2001, section 3, p. 1.

[33] Rochet and Tirole, "Platform Competition in Two-Sided Markets," pp. 23–24.

[34] Jon Friedman and John Meehan, *House of Cards: Inside the Troubled Empire of American Express* (New York: Kensington Publishing, 1992), pp. 13, 56.

accounts were "marquee" customers, who allowed American Express to raise its prices to the other side of the market, merchants. Similarly, marquee customers—in the guise of popular stores, often called anchor tenants—are important for shopping malls as well: by attracting customers they make a mall more attractive to other stores. The decline of a marquee store can sound the death knell for an entire mall.

In the software world, marquee customers are usually businesses on the user side and "killer applications"—an application so innovative and popular that people and businesses buy the computer system mainly because they want the app—on the developer side. VisiCalc was the killer app for the Apple II computer. It was one of the most important reasons behind the initial popularity of this platform. Likewise, Mario Bros. was largely responsible for Nintendo's millions of sales of its NES video game console in the United States, and Sonic the Hedgehog was the main reason for its displacement by Sega's Genesis as the dominant console several years later.

Multihoming As Table 2 illustrates, customers on at least one side of a multisided market often belong to several different networks. This is known as multihoming. Take payment cards. Most merchants accept charge, credit, and debit cards associated with several systems; consider how many card symbols there are at the next gasoline pump you use. On the other side of the market, the average consumer has 3.6 payment cards.[35] Advertisers typically place advertisements in several different magazines, and consumers read various magazines.

In general, multihoming by one side of the market relaxes platform competition for that side and intensifies it on the other. For instance, if game developers suddenly become more prone to porting their games to both Sony PlayStation 3 and Microsoft Xbox 360, there would be less reason for Sony and Microsoft to hold royalties down to attract developers. Moreover, in this case the two consoles would become closer substitutes from the users' perspective, since they would have more games in common, so one might expect the battle for the end users (many of whom buy only one console) to become fiercer, resulting in lower console prices.

Sometimes unrelated platforms evolve into intersecting ones, which target one or more groups of customers in common; we will see this for digital media platforms. Platform competition can be fierce when either group of customers is price-sensitive because they have other alternatives. The *Houston Chronicle* may have 89 percent of the newspaper readers in Houston, but that does not mean that it can exercise a great deal of pricing power.[36] Advertisers have many other ways of getting messages to readers, so they are sensitive to prices. And while readers may not have many newspaper alternatives, they do have other ways of getting the news, and having a lot of readers is what makes advertisers pay.

35. http://www.cardweb.com/cardtrak/pastissues/april2004.html.

36. This number is the total daily circulation of the *Houston Chronicle* divided by the total daily circulation of all daily newspapers in the Houston area. *Circulation* 2003, SRDS (2002), p. 67.

Table 2 The Presence of Multihoming in Selected Multisided Platforms

Multisided Platform	Sides	Presence of Multihoming
U.S. residential property brokerage	• Buyer • Seller	*Uncommon:* Multihoming may be unnecessary, since a multiple listing service allows the listed property to be seen by all member agencies' customers and agents.
Securities brokerage	• Buyer • Seller	*Common:* The average securities brokerage client has accounts at three firms. Note that clients can be either buyers or sellers, or both.
Newspapers and magazines	• Reader • Advertiser	*Common:* In 1996, the average number of magazine issues read per person per month was 12.3. Also common for advertisers: for example, on August 26, 2003, AT&T Wireless advertised in the New York Times, (the) *Wall Street Journal,* and the *Chicago Tribune*, among many other newspapers.
Network television	• Viewer • Advertiser	*Common:* For example, viewers in Boston, Chicago, Los Angeles, and Houston, among other major metropolitan areas, have access to at least four main network television channels: ABC, CBS, FOX, and NBC. Also common for advertisers: for example, Sprint places television advertisements on ABC, CBS, FOX, and NBC.
Operating system	• End user • Application developer	*Uncommon for users:* Individuals typically use only one operating system. *Common for developers:* As noted earlier, the number of developers that develop for various operating systems indicates that developers engage in significant multihoming.
Video game console	• Game player • Game developer	*Varies for players:* The average household (that owns at least one console) owns 1.4 consoles. *Common for developers:* For example, in 2003, Electronic Arts, a game developer, developed for the Nintendo, Microsoft, and Sony platforms.
Payment card	• Cardholder • Merchant	*Common:* Most American Express cardholders also carry at least one Visa or MasterCard. In addition, American Express cardholders can use Visa and MasterCard at almost all places that take American Express.

Scaling Many successful multisided firms seem to have adopted a fairly gradual entry strategy in which they scale up their platform over time.[37] Many payment card systems, for example, started in one city or region before expanding nationally. It is often difficult to predict just what the right technology and operations infrastructure will be. Therefore, the multisided firm may find it advantageous to establish efficient buyer-seller transactions and balanced pricing first, and make large investments only after the platform has been tested. Platforms such as eBay, Palm, and Yahoo! have expanded gradually and methodically, building up customers on both sides of their markets.

Strategy Markets hardly ever cooperate with professors by following simple textbook rules exactly. But in traditional markets there are classic truisms that can at least serve as a benchmark, a starting point for more nuanced analysis. By contrast, multisided platforms, especially those in new markets, all too often require clean-sheet planning. With multiple yet interdependent business constituencies to serve, costs provide little guidance for pricing strategies. By the same token, early entry may yield first-mover advantages or provide an instructive failure that simplifies the search for successful strategies by businesses that follow. And, in light of the interdependence between different stakeholders, changes in the business environment may have multisided effects that are very difficult to anticipate.

III. OPEN-SOURCE SOFTWARE

The fortune made by the founders of Google is built in part on the efforts of thousands of volunteers around the world who helped develop the operating system that powers the massive array of server computers that helps us conduct searches and in return peppers us with customized advertising messages. Google uses Linux. Like anyone else, it can download this software platform for free from the Internet and customize the source code to meet its own needs. As of 2004, around 19 percent of server computers ran Linux worldwide.[38]

This is almost unthinkable in any other industry. It could not happen in manufacturing, because someone would have to pay for the raw materials to assemble an automobile, for example. Yet even in intellectual property-based businesses such as movies one seldom sees products made by volunteers beating those built by for-profit firms. Linux is the result of the remarkable open-source production model. We turn now to the underlying economics of that model.

Open source is based on a decentralized method for producing software that relies heavily on the Internet. Programmers working on their own or through their companies contribute code to

---

37. An example of a failed strategy is the case of Chemdex, a business-tobusiness marketplace, and its parent company, Ventro, which made initial technological and operational investments in the hundreds of millions of dollars (http://www.zdnet.com.au/newstech/enterprise/story/0,2000025001,20107754,0 0.htm).

38 Al Gillen and Dan Kusnetzky, "Worldwide Client and Server Operating Environments 2004–2008 Forecast: Microsoft Consolidates Its Grip" (ID C report no. 32452), December 2004, table 2.

open-source projects. The source code of the resulting programs is made available for free; hence the term open source. Users must sign a license that requires that if they redistribute the program, even in modified form, they must make the source code available. As a result, it is hard to make money directly from open-source programs or anything derived from open-source programs. Open source began as an ideology—"free software is a matter of liberty," according to Richard Stallman—but has evolved into a multi-billion-dollar business based on selling hardware, software, and services that are complementary to open-source programs.[39]

A. The Production of Open-Source Software

In its early days, individuals who donated their time to work on projects that interested them were the main contributors to open-source software. Typically, a person or a small group of people gets an idea for a project that is interesting, useful, or both. The original developers begin work on the project and eventually solicit support from other interested programmers. Over the course of the project, programmers, including the original developers, may come and go as they complete work and as their interest waxes or wanes.

The programmers communicate with each other over the Internet. A core group, often consisting of one or more of the original developers, has responsibility for incorporating changes and suggesting things that need to be done. Modified versions of the source code are posted on the Internet and available for free to anyone who wants to use them or modify them further. Over time, users regularly identify bugs that had originally escaped detection, and worthwhile features to add. These users can provide feedback to the developers (or become developers themselves). Through this ongoing process the software becomes tested, debugged, and developed.

The Apache Web server is one of the most successful and famous open-source projects. An early version was written at the National Center for Supercomputing Applications (NCSA) and became the most popular Web server by 1995. Development stalled when Rob McCool, the core developer, left NCSA. Following his departure, some Webmasters began coordinating their fixes via email. Eventually, the Apache group, consisting of eight core contributors, was formed. In April 1995 the first version of the Apache server (version 0.6.2) was released, and became a huge success. The server was completely revamped during the second half of 1995, and Apache 1.0 was released in December 1995. Less than a year after its release, Apache 1.0 became the most popular Web server in the world.[40] The Apache group was incorporated in June 1999 as the Apache Software Foundation. Apache 2.0 was released in 2002, and minor fixes and updates have been periodically released since then. Apache remains the most popular Web server in use, with more than a 50 percent share of its segment.[41]

---

[39] http://www.fsf.org/philosophy/free-sw.html.
[40] http://httpd.apache.org/ABOUT_APACHE.html.
[41] http://en.wikipedia.org/wiki/Apache_web_server.

This production method differs from the commercial approach.

First, there is typically little analysis of consumer needs other than introspection: "What would I like my software to do?" This may be augmented by user feedback, but these users are self-selected; except in unusual circumstances, they are not drawn randomly from the universe of potential users of the software.

Second, there is little formal testing of the type that commercial firms often must engage in: internal testing using hundreds, perhaps thousands of hardware and software configurations in a controlled manner. Testing is instead performed by the users who try versions of the software in uncontrolled environments, much like beta tests for commercial software developers (although perhaps with more sophisticated users providing feedback to the developers).

Third, the development of open-source software is less structured than the development of proprietary software. Although the core developers may provide direction, changes in the software result much more from individual action.

As open source has evolved, commercial businesses have become more intimately involved in steering open-source projects. They do this by having their employees spend time contributing open-source code and working on the various committees that oversee open-source projects. In a 2003 survey of open-source contributors, nearly 15 percent reported that their employer paid them to develop open-source code, 13 percent noted they were paid to "support" open source, and 13 percent stated they were paid to "administer" open-source projects.[42]

IBM is arguably the best example of a traditional for-profit company with strong ties to open-source software. The bond was officially created in 2000, when IBM announced a $1 billion investment (including marketing expenditures) in a variety of open-source initiatives, including adapting Linux and Apache to IBM's various computer hardware platforms.[43] IBM's hardware business was unusual in that it marketed several fundamentally different types of servers with mutually incompatible operating systems. Adopting Linux permitted IBM to unify its server product line, so that proprietary IBM software (and other software) could be used on all the different servers. By making Linux available on all of its servers, from the smallest to the largest, IBM added consistency to its product line that was missing before. IBM therefore had an incentive to do open-source development that would make Linux run (or run better) on its servers because the investment would provide benefits to IBM.[44]

The open-source investment strategy appears to have paid off handsomely for IBM. For example, China's postal service hired IBM to build Linux-based networks for over 1,200 of its branch

---

[42] http://www.idei.fr/doc/conf/sic/papers_2005/pdavid_slides.pdf.

[43] Joe Wilcox, "IBM to Spend $1 Billion on Linux in 2001," *CNETNews.com*, December 12, 2000 (http://news.com.com/2100-1001-249750.html).

[44] James Evans, "IBM to Invest Almost $1 Billion on Linux Development," InfoWorld, December 12, 2000 (http://www.infoworld.com/articles/hn/xml/00/ 12/12/001212hnibmlin.html).

offices.[45] A year after its initial $1 billion investment, the company announced that it had already recouped that amount and more.

B. Intellectual Property Rights

The proponents of open-source software faced a problem. On the one hand they wanted to make open-source software widely available. That meant that they did not want to use copyrights, patents, or trade secrets to limit the distribution of open-source programs. On the other hand, they wanted to make sure that commercial enterprises could not free-ride on the efforts of the open-source community by making minor changes or additions to open-source programs but then enforcing their own intellectual property rights on the entire modified programs.

The General Public License (GPL) was an ingenious solution to this dilemma. The GPL is based on "copyleft":

> You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.[46]

(Despite the copyleft name, the GPL is enforced by copyright law. Copyright is the source of the property protection that enables those who release software under a GPL to impose conditions on others who obtain that code.) The copyleft provision means that if people choose to distribute software that is based in part on other software covered by the GPL, they must distribute their new software under the GPL. GPL software thereby propagates itself. Copyleft makes it difficult for anyone to earn significant profits from selling software code subject to the GPL. As Richard Stallman observed,

> We encourage two-way cooperation by rejecting parasites: whoever wishes to copy parts of our software into his program must let us use parts of that program in our programs. Nobody is forced to join our club, but those who wish to participate must offer us the same cooperation they receive from us.[47]

Proprietary programs can use or communicate with GPL programs in some limited ways without themselves becoming subject to the viral license condition, but the FSF recognizes that the dividing line can be murky. The terms of the GPL apply only to the distribution of software licensed

[45] http://www.infoworld.com/articles/hn/xml/03/01/23/030123hnibmlinux. html?0124fram.
[46] http://www.fsf.org/licensing/licenses/gpl.html.
[47]. http://www.rons.net.cn/RMS/ms_oss.html.

under the GPL, although what "distribution" means in this context is not entirely clear either. It may be possible for an enterprise to modify a GPL program and use it internally without being legally bound to make the source code for its modified version available to others. On the other hand, if the same enterprise distributed its modified GPL program to a subsidiary, the terms of the GPL might well require it to make the source code available to all comers.

Most open-source projects are subject to the GPL. However, several commercial ventures have chosen to use modified licenses. The two most prominent examples are the Common Development and Distribution License (CDDL) that covered Sun's Solaris as it went open source and the Mozilla Public License (MPL) that governs the Firefox browser, among other products. Both contain provisions that GPL does not, and thus code cannot be freely moved between GPL and projects covered under these other licenses. Opponents of this balkanization of open-source licenses contend that it leads to islands of legally incompatible code. For example, owing to different licenses, no cross-pollination between Linux (GPL) and Solaris (CDDL) is possible. Proponents argue that companies have varying needs and catering to these differences is necessary for open-source software to flourish. In addition to relying on more restrictive licenses some open-source software companies are using intellectual property rights to help protect their investments and guard their profits. Red Hat, for example, has used trademark law to help protect its compilation of Linux from others.

C. Incentives

The incentives for writing open-source software are different from those for writing commercial software. Many people write open-source code without being paid directly for it. These are volunteers who write code in their spare time because it interests them. Others write code because their companies have asked them to. This may sound traditional but is not, since their employers cannot sell the resulting code or obtain intellectual property rights over it.

Why Individuals Work on Open-Source Software Why programmers donate time to open-source software projects is a subject that has generated considerable discussion.[48] Open-source advocates have suggested several motives, four of which involve nonfinancial rewards:

- It is a good way to learn how to program and develop skills.
- It is fun. Since a programmer is free to pick and choose among open-source projects, he need only work on matters of interest.

48. For a more detailed discussion of open-source software development, see Joshua Lerner and Jean Tirole, "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 2002; Joshua Lerner and Jean Tirole, "The Open Source Movement: Key Research Questions," *European Economic Review* 45, nos. 4–6 (2001): 819–826.

- It is prestigious. Success at open-source development rates highly among those whose opinions most programmers most value—other programmers.
- It "scratches an itch." Programmers attack problems that they personally face or because they are intrigued by the intellectual challenge.
- It meets an ideological urge—the desire for free software and the "liberty" it entails.

The "scratches an itch" motive has been considered by some analysts as leading to something like a cooperative of users. A number of developers all consider a particular type of software potentially useful, so they pool their talents to develop the software. With this type of motivation, the GPL has sometimes been considered beneficial as an enforcement mechanism: it ensures that no one can take the collective intellectual property, add some private intellectual property, and treat the whole as a private good.

Business Models Based on the Open-Source Concept Businesses have incentives to "donate" employees to the development of open-source projects that stimulate the demand for other products or services sold by the firm. This has become an increasingly large source of labor for open-source projects.[49] IBM and Red Hat illustrate the motivations. As discussed earlier, IBM's model is built on driving the sales of its key products: supporting Linux software increases IBM's sales of hardware, proprietary software, and services. Linux offered a way for IBM to integrate its entire line of servers without having to develop a software platform of its own, and without having to shoulder the continued support and development of that system on its own.

Red Hat is a somewhat different story. That company began as a pure open-source vendor offering a distribution of Linux. Over time, it has gradually moved toward more traditional software licensing, presumably because it is difficult to support a for-profit company with a pure open-source business model. Red Hat is focused on solving a problem inherent in the Linux development model. For major proprietary operating systems such as Windows, the components of the software are integrated by the distributor and sold as a single program. Since no one developer exists for Linux, bits and pieces of the operating system tend to float around—some in forms unusable by nonprofessionals. Specific Linux distributions consolidate these bits and pieces into a convenient package.

Red Hat is arguably the premiere Linux distribution, with more than 46 percent of Linux server distribution shipments in 2004.[50] The company was founded in 1995 and subsequently

[49] "OpenOffice Team Wants IBM Contribution," *VNUNet*, April 25, 2005; Timothy Prickett Morgan, "Novell Creates Hula Open Source Collaboration Server," *ComputerWire*, February 16, 2005 (http://www.computerwire.com/ industries/research/?pid=23554158-E49D-4ED2-9482-72B4B5D4119F&type= CW%20News).

[50] Al Gillen, Milla Kantcheva, and Dan Kusnetzky, "Worldwide Linux Operating Environments 2005–2009 Forecast and Analysis: Product Transitions Continue" (IDC report no. 34390), table 2.

enjoyed significant growth, topped off with an IPO in 1999 that generated the eighth biggest first-day percentage gain in Wall Street history.[51] Like many other high-tech companies, Red Hat lost quite a bit of value in the dot-com crash, but it has since rebounded successfully.

Red Hat really does three things. First, it integrates components of Linux into a cohesive distribution, including commonly used open-source products along with the core operating system. Second, it adds its own software to provide a better user experience and to make installation and updating easier. Third, it sells support packages and certifies that external administrators are qualified to work on Red Hat products. The company includes only open-source software, and the code it writes is licensed mainly under the GPL.

Red Hat changed its business model drastically in 2003 by splitting its distribution into two products—the Fedora Project, a more traditional open-source project, and Red Hat Enterprise Linux (RHEL), the flagship product. Fedora is the place for experiments to run and outside developers to submit code, while RHEL is a stable version of Linux for paying customers. Along with the split came a new licensing agreement. RHEL source code is available for free from Red Hat, but the code computers need to run the operating system is available only with the purchase of a support subscription.[52] And support subscriptions must be purchased for each computer, just like traditional proprietary software licenses.

One fundamental problem with generating revenue from GPL software is that anyone can take the source code, "compile" it for computers to read, and resell it without incurring the original creator's development costs. Red Hat has tried to sidestep this problem. Another company could rebuild RHEL from freely available source code, but it would have to strip out all references to Red Hat to comply with trademark law. Purchasers could not be certain that the distribution really contained all of the pieces in Red Hat's, or that the installation would work as seamlessly. Thus, Red Hat has used its reputation in combination with trademark law to limit the potential for another company to undercut its profits.

Open source seems so New Age. Yet when one looks over the history of the computer industry, it turns out that the business of selling software—including software platforms—really didn't take hold until the late 1970s. Microprocessors created a mass market for software that attracted entrepreneurs. Many of these pioneers wrote applications that would run on the new microprocessor-based personal computers. A few focused on refining software platforms whose shared code could be used by many developers and customers at the same time.

[51] W. G. Rohm, "Inside The Red Hat IPO," *Linux Magazine*, November 15, 1999 (http://www.linux-mag.com/1999-11/redhatipo_01.html).

[52] Advanced users could compile the source code themselves. As discussed later in the chapter, this has many disadvantages for corporate customers.

INSIGHTS

- Like other information goods, software platforms are produced by educated workers, are malleable and easily changed, and are reproducible at virtually no cost.
- Bundling features into the software platform is often efficient for the platform producer and for end users, as it is for most information goods, because it lowers distribution costs and expands demand.
- Software platforms create value by reducing the costs for their multiple customer groups to come together and thereby enhance the value that each customer group delivers to the other. They do this mainly through providing shared services—made available through APIs—that reduce costs for developers and users.
- Multisided platforms must consider marginal costs and price sensitivity in pricing, like single-sided businesses, but they must also consider which side values the other side more. Software platforms generally charge low prices on one side in order to attract customers who can then be made available to the other side. Getting the balance right among all sides is more important than building market share.
- Commercial and open-source production methods have both proved viable models for producing software platforms. Commercial methods seem better suited for managing the multisided aspects of platforms, while open-source methods have produced reliable platforms and applications.

CHAPTER

16

What's Next in Payments: Invisible Engines

David S. Evans

ABSTRACT

The use of software platforms to drive innovation and transform industries has exploded since the 2006 publication of my book Invisible Engines *with MIT Professor and former Sloan School Dean Richard Schmalensee and Harvard Business School Professor Andrei Hagiu. Around the globe, invisible engines are ushering in a new era of technological change based on software. The Apple iPhone has shaken the mobile phone industry worldwide in part by creating a massive applications business built on the phone's operating system. Firefox has revolutionized the browser industry by encouraging developers to write add-ons and, in doing so, toppled Microsoft's Internet Explorer from dominance in many countries. Facebook has created a powerful social networking platform by opening itself up to developers. Amazon started a cloud-computing platform that enables entrepreneurs to access its vast software code, hardware and global communication systems over the Internet.*

This chapter explain how software platforms are going ready to transform the payments industry and why you will be touched by this new revolution and perhaps even want to be one of the entrepreneurs or investors who pioneer this new area. I'm first going to provide background on the technology and business models behind invisible engines. I'm then going to show why this is important for payments and describe some of the main players that are leading the revolution at this point.

I. INVISIBLE ENGINES AND THE APIs THE DRIVE INNOVATION

Invisible engines are based on software code. Software programs are written in various languages. Working by themselves or with other programs they tell hardware – from the pixels on your iPhone to the chip on your desktop to the communication devices providing your internet connection – what to do. They are the brains behind everything from the card reader in which you swipe your debit card, to your social network page where you visit friends, to the complex trading decisions at hedge funds, to your favorite word processing package.

Software programs can become platforms that support other software programs when code that is written to perform a particular task is made available to other software programs so the writer does not have to write that code again. That is what Windows does. It contains many little bundles of code that help developers write their own programs. Microsoft provides a link to that code—called Application Programming Interface (API)—that allows developers to link into the Windows code. Some businesses make their service available by providing an API. For example, YouTube provided an easy-to-use API for people to include access to videos on their social network pages.

By making APIs available, the owners of software programs are letting others use their intellectual property, the results of their hard work writing code. They can benefit from doing this if others decide to write valuable applications that work with their software platform. More people are likely to use a platform if there are more attractive applications that work with it. For example, many users decided to switch from Internet Explorer to Firefox because there were add-ons for Firefox that did things that were not available with Internet Explorer such as support for XHTML – a popular programming language for webpage developers. Software platform providers do not have to make their code available for free though. Developers that sell their applications through the iPhone app store give a portion of their revenue (about 30 percent as of the end of 2009) to Apple.

Invisible engines are catalysts that use a multi-sided platform business model. They ultimately create value by making it easier—lowering the transactions cost—to get several different groups, whose members value each other, together. In the physical world, a shopping mall is a two-sided platform that helps bring shoppers and merchants together.

Software platforms generate value by reducing the cost to developers of writing applications for consumers and providing a common place for users and developers to get together. Facebook for example provides a place, not only for friends to get together, but also a place for friends to find applications that can help them and applications to find users that can benefit from them. Facebook helps developers do this and thereby creates value for its community and ultimately itself.

II. HOW THE IPHONE INVISIBLE ENGINE PROVIDED A CATALYST FOR THE MOBILE PHONE INDUSTRY

The iPhone provides a potent illustration of the invisible engines/catalyst business model. By the end of 2009 there were more than 100,000 applications available for the iPhone. These applications have been downloaded more than 3 billion times since Apple opened its iPhone store in July 2008. Apple's success has shaken up the mobile phone industry in many countries around the world. It has unleashed tens of thousands of entrepreneurs and created a multi-billion dollar application industry for mobile devices. Here is how an invisible engine helped make this.

Inside the iPhone there is a computer chip on which is stored the software code that does all the magic that has been behind the success of this smart phone. That software code is usually called the operating system for the iPhone. Apple could have decided to keep that code entirely to itself and focus on making the iPhone a self-contained device that only did things that Apple wanted it to do; it would have had a "closed software platform". Instead, in a smart move, it decided to make that operating system a software platform that could support third-party applications. That meant it had to create APIs which would enable outsiders to link into the iPhone's operating system. These APIs give developers access to blocks of code that manipulate various aspects of the iPhone. Apple then provided a "software developer kit" (SDK) to help programmers develop applications using the iPhone's APIs and put together a web site to provide other resources for developers.

An important set of business dynamics starts when such an operating system is opened up and thereby becomes an "open software platform" that can support developers. Apple initially provided more than enough applications and features on the iPhone to persuade many people to get one. By opening up its operating system it then started a virtuous circle. Entrepreneurs and hobbyists saw the opportunity to write interesting applications that could use the features of the iPhone, including its large touch screen and connectivity. The iPhone users soon found that they could download applications that made their phones even more useful than they were when they first bought them. More people started buying iPhones because of the increasingly valuable applications. As the number of iPhone users increased, developers had even more incentives to write applications.

Apple has not taken this process for granted. Like most other successful platforms it has stoked the fires. It developed a marketplace--an online shopping mall--to help developers make their applications available to iPhone users and earn revenues from them. Apple is responsible for collecting money from iPhone users and then gives the developers about 70% share of the revenue. Users have benefitted from having effectively a single location to search for applications. Developers have benefited because Apple has reduced the transactions costs of selling and has assembled an audience of users.

Apple has obviously made enormous profits and seen the market value of the firm increase because of the success of the iPhone and its application store. iPhone users have done very well also as a result of the tens of thousands of applications they can choose from, some of which enable them to do things that were not possible before. New markets have opened for entrepreneurs that use the iPhone, and some of them are earning significant profits or obtaining fame. By bringing developers and users together the iPhone has served as a catalyst which has created value—and profit—out of thin air. Others have followed suit. The most significant follower is Google, which has introduced the Android software platform for mobile phones, has stimulated hardware makers to introduce phones using this platform, and has worked hard at persuading developers to write applications for its store. (Microsoft has been trying to be the open software platform for mobile for many years but has thus far failed to catch on.)

Apple's experience with the iPhone shows the power of the invisible engine model. As we will see software platforms can drive payments innovation and revolutionize the payments business.

III. THE NEW AGE OF INVISIBLE ENGINES: HOW SOFTWARE PLATFORMS WILL DRIVE GROWTH IN THE NEXT DECADE

The invisible engine model followed by Apple's iPhone has helped power the information technology industry for about three decades. Apple itself was one of the pioneers in encouraging developers to write applications for its desktop operating system. It invented the "software evangelist". Microsoft, though, was the maestro catalyst. Its Windows software platform attracted thousands of software developers and hundreds of millions of users. As a result it has been the durable center of a vast desktop-computer based ecosystem since at least the launch of Windows 3.0 in 1990. My book, with Andrei Hagiu and Richard Schmalensee, Invisible Engines tells the story of how software platforms have transformed industries including computers, video games, and handheld devices and then accelerated innovation. It then examines the forces behind the various business models that have been adopted in these industries. Many software platforms have decided not to charge developers while others have. Some have decided to vertically integrate into hardware while others have remained pure platforms.

The development of the internet and the spread of high-speed broadband throughout the world provided have resulted in a new age of software platforms. Several of the major web businesses have turned themselves into platforms. The heart and soul of any web business does not reside in a server farm somewhere or in its buildings—it lies in the thousands and thousands of lines of software code that enable people to see and interact with web pages. Once written these software programs can be opened up to others by exposing APIs that enable developers to use portions of that code to interact the web properties.

Facebook has allowed developers to write applications that work inside its social network. As a result developers can write games, such as Farmville, that people on Facebook can play or online shopping and advertising programs such as those developed by FreeCause for raising money for charities. By the end of 2009 there were more than 500,000 (yes, really) applications running on Facebook. This five-year old site has also developed a program called Facebook Connect. Developers can write applications that through APIs made available by Facebook enable users of those applications to pull their Facebook identity, friends and privacy to other websites. As of the end of 2009 there were more than 6000 developers working on Facebook Connect applications. These applications make Facebook more valuable to its users, and the greater number of users makes Facebook more valuable to developers. Although it has made slow progress, the expectation is that Facebook will make money by selling, in some fashion, access to the eyeballs drawn to Facebook and these applications to advertisers.

Amazon is one of several companies that are competing to create a general purpose software platform that can be accessed over the internet. Such platforms are said to reside "in the cloud" (more on this below) because they are off in the distance as opposed to on the user's desktop or other client computer. The idea is to develop a set of services that application developers can use. These include software services that allow developers to use the code rather than writing their own.

But they also involve access to the vast server farms and communication networks around the globe. Amazon's Web Services plays to this company's strength in selling products over the internet. It provides special services to developers who are writing programs to help merchants operate virtual stores and receive payment.

Invisible engines are beginning to disrupt the payments business both online and offline. This industry is at the heart of commerce and makes it possible for businesses and people to exchange value around the world. By definition it involves essentially all the money in the world. If invisible engines start driving innovation in this industry they could have an enormous impact on consumers and businesses who transact with each other.

IV. WHAT'S IN THE CLOUD FOR PAYMENTS?

When we speak of "payments in the cloud" we are using the term in a broader sense than it is usually used in the information technology world.

Cloud computing simply refers to accessing software and hardware over the internet. In a sense we've all been doing that since the commercial internet started 15 years or so ago. Just consider search. Google stores massive amounts of data on web pages on server farms and updates these data regularly. It has devised complex and sophisticated software that searches those web pages for information. We obtain access to that software and hardware using just our browsers (and possibly some code that is stored on our computers if we have installed a toolbar).

As the web economy has grown, more businesses have started offering services that reside in the cloud. To take a few examples, Flickr helps consumers store and manage your photos in the cloud; Salesforce.com provides customer relationship management software in the cloud. Google Docs helps consumers and enterprises do word processing, spreadsheets, and other applications in the cloud and Microsoft's soon to be released Office 10 will do so as well.

Several things have happened that have led to cloud computing becoming the next big thing.

1. Accessing things over the internet has become much faster and more reliable in many parts of the world as a result of the spread of broadband.

2. Software development has moved from local devices to the internet so there are more languages, tools, and code to leverage.

3. There's been a lot of progress in developing inexpensive, efficient, and massively large computer systems. Of necessity Google and other companies had to figure these kinds of things out.

4. The companies that have invested in massive and scalable server farms have realized that they can sell others access to it. Many of the obvious suspects have gotten into offering cloud-based services including Amazon, Google, IBM, and Microsoft.

Selling on Your Own Website

To get your own branded, custom online store backed by the same technology that powers Amazon.com, sign up for WebStore by Amazon or try it for free for 30 days. WebStore by Amazon can provide a simple, secure and reliable solution to your online business needs. Preview a store created in a matter of minutes using 1-Click WebStore.

If you need options for collecting payment for sales on your own website, Amazon Payments has several programs just right for you. Checkout by Amazon is a complete checkout solution that provides your customers with the same secure and trusted checkout experience available on Amazon.com. Amazon Simple Pay is a set of payment-only products that allow your customers to use payment information from their Amazon.com account for orders on your website. Amazon Flexible Payment Service is a set of APIs for developers that allow the movement of money between any two entities, humans or computers.

Drive traffic to your website and make Amazon.com customers your customers. Amazon Product Ads is a cost per click advertising service. Simply upload your product catalog and set your bids and budget to get your ads live on Amazon.com. We'll show your products in highly targeted placements in search and browse results and on product detail pages. Customers click through to your website to purchase directly from you. Learn more or Sign up today!

Cloud computing is also associated with a change in the traditional business model. Consumers and enterprises typically bought (technically they licensed, but usually in perpetuity) software packages which they installed on their computers. Expanding information-technology resources meant front capital expenditures for hardware and software. In contrast, cloud computing is usually sold in bite sizes under what is known as the "software as a service" model. In fact it is even better than that. In a typical situation consumers and enterprises are paying monthly or annual fees to obtain access to hardware including storage as well as software.

Amazon Merchant Services provides a good example of what cloud computing is all about and one that is relevant for the payments industry. Suppose you are a retailer who wants to start selling over the internet. If you wanted to stay on earth, you would hire someone to build an e-commerce site for you, line up someone to host it, and get a processor to enable you to take payment cards online. If you wanted to go to the clouds, you could rely on the software that Amazon has created for quickly building a store on line, have them host it at one of their server farms, and use their payments system.

Amazon Merchant Services also illustrates how cloud computing has resulting in ***innovation moving to the cloud***. Amazon has developed a payment services product that relies on the payment systems back down on the ground. They connect to one or more processors that in turn connect to the rest of the payments system. Amazon's innovative product stays in the cloud where everyone down on earth can connect to it. PayPal, Apple, and other companies are also engaging in this sort of innovation in the cloud as we will see.

Cloud computing also unleashes innovation from hardware. For most of its history the payments business has been tied to devices that pretty much have the software hard-wired in. Maybe the supplier of the hardware could modify the software but the merchant really couldn't nor could other players in the ecosystem. And since these devices were closed systems a clever entrepreneur couldn't develop applications that could provide extra value to the merchant or to members of the payments value-chain. That can and will change. Software will come down to the devices from the cloud and APIs will enable entrepreneur/developers to add value.

There are other opportunities for moving innovation off the ground in payments that don't necessarily involve using internet connections to software residing on remote hardware. At the moment the payments system consists of a number of highly secure private networks that companies can plug into. A point of sale terminal at a retailer plugs into several private networks. It does that by connecting to a switch which then routes transactions across the debit or credit rails to the relevant clearing and settlement system and so forth. One doesn't have to rely on the internet necessarily to move innovation off the ground. Companies could provide remote services over private networks.

One way to do this would involve developing a software platform that enables third parties to write applications that would work with the payments system. That could happen in a couple of ways.

1. The major processors currently operate software platforms that provide services to banks. They could potentially open up portions of these platforms to developers to build applications that use some of the features and functionality of these platforms. The same is true for networks. Obviously, there are security issues that would have to be dealt with to ensure the safety and integrity of the platforms.

2. A third party could build an uber-software platform that sits on top of all of the various software programs that lie on the payment rails. That uber-software platform would connect into these software programs. Then developers could plug into the uber-software platform to write applications. In both of these cases it becomes possible for innovation to take place outside of the software and hardware that currently comprises the payments systems. The resulting innovative applications then also sit outside.

When we talk about payments in the cloud for the rest of this chapter we will use this term to refer to the development of applications that sit outside of the traditional payments system and link to the traditional payments system through APIs on software programs that sit somewhere on the rails of the traditional system.

One thing is for sure in my view: innovation in payments will move to the cloud and anything who wants to be a player will need to be there as well or work with someone or is.

V. WHY THE PAYMENTS INDUSTRY NEEDS A CATALYST TO DRIVE PAYMENTS INNOVATION

I've explored how invisible engines drive innovation and transform industries. I'm now going to explain how they will do that in the payments industry. This entry explains why the payments industry needs a catalyst to ignite innovation.

There is an intricate set of systems and enormous group of businesses that are involved in payments. They range from the companies that print paper checks, to ones that make the point-of-sale terminals where you swipe your debit card, to banks that operate depository accounts, to clearing and settlement networks. At the beginning of 2010 this industry involves many different computer systems, each with a software program, that have been through brute force forced to interoperate with each other. Consider what happens when you swipe your card. The terminal sends your card information along with details of the transaction to a switch. That switch, which may be operated by a variety of payment players, has a software program that decides what to do with the transaction. It will send it on to a merchant processor that will keep track of the details of the transaction and act as an intermediary with a clearing and settlement system which will contact the bank processor that acts on behalf of the bank that issued the card or possibly the bank itself. Many software programs running on numerous pieces of computer hardware are involved in the various steps in the process.

What we have just described is one "rail" in the payments system—the one for credit cards. While this may vary by country, in the United States there are other rails for debit cards, paper checks, electronic funds transfer, and other "tender" types. Now suppose that you are an entrepreneur who has come up with a great idea that involves incorporating payments into your application. This innovation could be something as simple as a software program that helps small businesses accept payment in multiple ways and integrate these payments into basic accounting software. To succeed, you would have to integrate your application into the many software platforms that are used for the many tender types taken by businesses. That is such a daunting task—one involving substantial money and time not to mention the cooperation of the other businesses that handle payments—that you might well give up.

This is a perfect problem for a catalyst to solve. Massive transactions costs make it hard for payments entrepreneurs and payments users to get together. A software platform—perhaps in the cloud-- can lower those costs by investing in linking to the multitude of software programs that handle various elements of payments. By exposing APIs, this software platform then makes it possible for entrepreneurs to quickly integrate into most relevant aspects of the payments business. Such a software platform is analogous to Windows which among other things enables software applications to work with the multitude of device drivers that work various hardware peripherals such as computer hardware, printers, and cameras.

A great deal of innovation can be unleashed once these APIs are exposed. Entrepreneurs won't need to replace hardware at the point of sale; they will just need to develop applications that work with platforms that connect to the hardware. They will be able to achieve massive distribution quickly thereby making it easier to ignite their businesses. With that they will be able to obtain scale more quickly and lower costs. It will be possible to roll out innovations to many channels simultaneously by porting apps between different platforms or using APIs for platforms to connect to multiple channels.

In the next section I will describe a platform that sits on top on the rails I've just described and helps developers of new payments businesses access the messy plumbing of the payments system right down to the point-of-sale devices at retailers. Then I will describe how PayPal has created a software development platform for online transactions. These platforms help developers of applications, say for the iPhone, easily integrate payments. And finally I will explain how the traditional card networks will play in this new world in which innovation, and value-added services, are driven by invisible engines Should the card networks create their own open software platforms or focus on just running the rails?

Innovation in payments will, in my view, be driven over the next decade by software platforms that facilitate a vast ecosystem of applications. And there's going to be competition over who provides the best APIs (and underlying code) to support applications. One of the things we've learned about software platforms is that they have huge network effects. The platform that gets more applications and users will get even more applications and users. That doesn't mean that only one software platform will survive because each could differentiate itself by catering to different needs of developers. But it does mean that it is unlikely that more than a handful of payment platforms will survive. The race is on.

VI. CAN IP COMMERCE CREATE THE APPS STORE FOR PAYMENTS?

There's no Steve Jobs strutting on the stage, massive consumer buzz, or slick television ads about how there's an app just for you. But there are some things you've probably heard about and a soon-to-be known Denver based company, IP Commerce, is the invisible engine that's powering them.

You might have seen the television advertisements for American Express OPEN. They've launched AcceptPay which they tout as "an online invoicing and payment solution that can help business owners improve cash flow at a time when customers are taking a longer time to pay." Amex's solution is based on a software application developed by PaySimple and a small-business service built around this application. A small business can easily accept multiple tender types online and offline after installing this application and signing up for the service. For the reasons discussed above, developing an application like this is incredibly complicated because it needs to work with multiple payments rails and comply with lots of security rules.

If PaySimple had to do all of that work itself, on top of developing easy-to-use software for small businesses with a greater user interface, it would still be a powerpoint deck gathering dust in a VCs office. Fortunately, PaySimple was able to use APIs which linked into IP Commerce's software platform, that provided all of the code and connections needed by link into the payments plumbing. Through the miracle of the division of labor, PaySimple got to focus on what it could do best—develop an innovative solution for small business—and offload a lot of dirty work to a specialist. As a result, American Express got access to a technology that would have taken it a long time and a ton of money to develop on its own—so much so that the AcceptPay innovation probably never would have launched without something like IP Commerce.

The APIs that IP Commerce makes available are enabling many entrepreneurs to develop payments-related applications far more quickly and cheaply than they could have done on their own or through other solutions. Square—the new smart-phone enabled payments system for small merchants and their consumers—is powered in part by the IP Commerce invisible engine. So is BrainTree Payment Solutions, a gateway and value add provider of online payment solutions. A new e-commerce application that relies on IP Commerce's APIs, getta , that just launched on MySpace. Getta uses the power of social networks to encourage group buying—the buyers get volume discounts and the merchants get sales.

IP Commerce has made these applications—and more like them—possible by spending years doing the dirty work of creating code that links into the multitude of discrete and non-interoperable software programs that ultimately run the payment systems. This tedious work is what many successful software platforms do. There's nothing glamorous about Windows 7 or Snow Leopard. These software platform—that we don't really think much about, except when they crash—are based on millions of lines of code that do all the nasty business of computing—from telling the chip to move 1s and 0s around, to working the pixels on your screen, to connecting to your home wireless system. IP Commerce has spent the last 4 years and many millions of dollars writing the code that's needed to get a lot of the mundane but complicated tasks that need to be done to work with each of the several tender types that are used in the US.

IP Commerce's work isn't done. While it covers much of the payments business through which most of the dollars flow it doesn't have 100% coverage yet. And it is going to need to keep up with ever changing software systems. Moreover, like all software platforms, its value will ultimately come from spawning numerous applications that rely on it. But it has done a lot of the tedious but important work of developing a cloud based platform that sits on top of the payments system and it has a rapidly growing stable of applications that have shown its power.

IP Commerce has the prospect of creating the same sort of powerful network effects that Windows has (which isn't to say they are necessarily the next Microsoft—that franchise is tough to beat!). These come in two ways. First, IP Commerce's platform will become more valuable to the participants in the payments ecosystem (such as processors and device makers) the more applications there are there are that drive volume through these points. Likewise it will become more valuable to developers of payments applications as even more of the payments system standardizes on this platform. Second, some of the applications that are being built on IP Commerce—like some of those built on Windows—have there own APIs and are creating there own ecosystems. These applications can work more easily together as a result of relying on the same platform.

Don't look, of course, for tens of thousands of applications coming out of garages over the next year. Creating "applications" in payments is a lot harder than it sounds. The success of the iPhone has made us think about applications as software programs that do cool things like help you find friends nearby or learn Mandarin. In payments, the software application is usually just one piece of a business model. Developing successful business models in payments—no matter how good the software application is—is extremely difficult. There's everything from glacial inertia to chicken and egg problems to solve.

Despite these challenges it is likely that cloud based platforms such as that provided by IP Commerce will become the one of the main drivers of innovation in payments in the coming years. These platforms drastically lower the cost of creating businesses that need to connect to the payments plumbing to create value. They are likely to therefore unleash a vast amount of entrepreneurial energy devoted to discovering ways to provide value to businesses, consumers, and governments that transact with each other. PaySimple, Square, Getta, BrainTree Payment Solutions, TIO Networks, SoftTouch, and over 100 other applications, are early indications of what's to come.

VII. PAYPAL X'S GLOBAL PAYMENTS DEVELOPMENT PLATFORM

When it comes to using invisible engines to drive innovation and transform industries, PayPal is the first out of the gate in the payments industry and well ahead of the traditional players. This online payments giant claims that "PayPal X is the first and only global payments platform open to third-party developers. Our new set of APIs will offer unlimited possibilities for [developers] to easily monetize your ideas, by providing security and connectivity to the world's financial systems." This isn't the usual hype—whizbangthing.ppt—that we get from payments providers trying to stake a claim. Six months after announcing its open platform more than 25,000 applications have been written to it and a vibrant developer community is cranking out more apps as you read this.

Here's what PayPal has done.

To begin with it turned PayPal into an open platform that provides a rich set of services to developers. It rolled out a new set of APIs that developers can use to build PayPal's payment services into their applications. Now, in addition to sending and receiving money for P2P and B2B transactions, developers can make parallel and chain payments, pre-approve with PIN-authorization, integrate PayPal into the merchant's order and account flow, pay attention to the context in which the merchant is interacting with the consumer, integrate ACH, pre-populate a PayPal account application for customers who don't have one, and many other things. These APIs link to code in PayPal's Adaptive Payments Platform that do the work for the developer.

Like every successful platform play, PayPal created a set of tools that help developers use these APIs. Just go to PayPal X Developer Network and you will see that PayPal has provided the standard "software developer kit" (sdk). Then it has provided a set of sample applications that give developers some guidance on what to do. PayPal X also have a "sandbox" where developers can test out their applications to debug them and make sure they work.

Taking a cue from the Invisible Engine strategy that Andrei Hagiu, Dick Schmalensee and I described in our book, PayPal put on a develop conference as soon as it opened its new platform. Over 1500 came to the November 3-4 conference in San Francisco. PayPal also offered prizes for the best applications. The winner of the $50k prize was Rentalic which helps people rent their stuff to other people. It has been described as a mashup of eBay and Craigslist. Here's how it works. Joe lists her lawnmower on www.rentalic.com. Sue offers to rent it for a fee they agree on. Joe uses PayPal to verify they can pay the deposit for it. Sue gets a secret code. After meeting Joe and making sure that she's getting a working lawnmower she gives Joe the secret code.

PayPal has been trying to get people excited about the iPad. On April 16th, 2010, it held a developer bootcamp in San Jose that was focused just on using PayPal X for iPad apps. It got 300 participants. Naveed Anwar, who is the head of Pay Pal X and seemingly its chief evangelists, talks about PayPal X and the IPad, and shows some clips of the developers beavering away on apps, in the Pay Pal X IPad Developer Camp video on YouTube.

PayPal is also rolling out its own applications but even here it is allowing developers to build value-added services on top of them. Most significantly, PayPal just rolled out the PayPal iPhone Bump which makes it possible for people to transfer money to each other using their iPhones. (Now, personally, on this one I'm a bit skeptical. The video provides a great example of overhyped P2P. Maybe there's a market for P2P, especially for small B to P transactions, but it isn't likely to be splitting meal bills which seems to be the favorite example of the P2P optimists, or charging a friend for a slice of my pizza as featured on the video.)

With X.com PayPal has ignited a new global payments platform very quickly. Here's the virtuous circle:

- They've started with rolling out a series of services that are available through APIs.
- They've poured money into encouraging developers to write applications by providing prizes and putting together a VC network.
- They've provided these developers with tools and soon a place where develops, like the iPhone Store, will be able to sell their applications.
- This has led to the rapid growth of applications, the growth of a developer community, and more buzz that will drive more developers to PayPalX.
- They've started building alliances with companies like Facebook, Microsoft, IBM, and Salesforce. That's the start of building out a potentially huge ecosystem.

It is remarkable how quickly PayPal has ignited its open platform strategy.

What does this mean for the traditional payment networks? For one it puts further distance between PayPal's online dominance and the traditional networks. PayPal conquered eBay, has made great strides in moving its payment services off eBay, and is now creating a potentially enormous ecosystem of online applications. This is coming at a time when the distinction between online and offline will soon end. For example, the iPad with wiFi at the point-of-sale is online and can compete with traditional hardware and private connections. PayPal and the traditional networks have had an uneasy relationship for the last decade. PayPal drives an enormous volume of card transactions but everyone knows that they have strong financial incentives to move transactions to ACH and that in any event PayPal—not a card network and not a bank—owns the online customer relationship.

For another, it means that the traditional networks are far behind their alternative payment system rival in seizing the most important opportunity of the coming decade: creating an open platform for development payments applications. First movers such as PayPal aren't necessarily the category winner—and in fact often get leapfrogged by those who bide their times. That said, PayPal—like Apple in the smart mobile space—is rapidly generating network effects and creating an increasing number of applications that will make it hard to beat.

Finally, PayPal's seemingly flawless execution of its open platform strategy makes one wonder whether the traditional card networks have the genes to compete in this very Web 3.0 kind of world. PayPal has its roots as a software application that lived within one of the great internet companies. It wasn't that long ago that the traditional card networks mainly used mainframes running Cobol. If the card networks can't pull it off though they could be to payments what the mobile carriers are to mobile communications.

VIII. THE CARD NETWORKS: STEALTHY TIGERS OR DEER IN THE HEADLIGHTS?

The traditional card networks are great at what they do and they've only gotten better over time. Transactions occur lickety-split over their networks. They provide valuable services to merchants and cardholders directly or through their bank partners. They have healthy market capitalizations—almost $180 billion for American Express, Discover, MasterCard and Visa combined (as of the end of April 2010). The old associations have made money faster than investors expected when they first IPO'd. Equity in MasterCard—the first of them to go public—sells for almost 5 times the June 2006 IPO price.

The card networks all recognize that their future lies in innovation. Hardly a day goes by without one of them (or several of them) announcing a new initiative somewhere in the world—from mobile to contactless to e-commerce. Indeed, it is the market's faith that these companies will seize the incredible opportunities for shifting transactions from paper to electronic methods that feeds the multiples they command (more than 30 for Visa at the end of April 2010).

It is unclear at this point, however, whether the card networks understand how open software payments platforms will radically change the payments business in the next decade. They may have lots of plans in the works, may be behaving stealthily, just waiting to pounce on the first movers in their own sweet time. Or they may be unsure of what's happening, what to do, and how to do it—they may be deer in the headlights of nimble rivals who aren't saddled with old code and pre-web says of doing things.

In fact, the card networks have three stark choices: spur innovation with an open platform, innovate with a closed platform, or just stick to running the rails.

- **Open software platform strategy.** They could use an invisible-engine strategy by building an open software platform on top of their existing networks and promoting the creation of applications that use services based on their rails. They would expose APIs to third party developers that would provide services. They would thereby compete for the attention of application developers with PayPal X, Amazon Merchant Services, IP Commerce, and whoever else tries to build a platform layer.
- **Closed software platform strategy.** They could try to develop value added services themselves in competition with third-party application developers writing for competing platforms, or partner with other firms to develop value-added services that rely on their network. With a closed strategy they would not provide APIs that third-party developers generally could access.
- **Just run the rails strategy.** They could focus on making money from transactions that go over their rails and focus on running those rails as efficiently as possible. They could leave operating software platforms and fostering third-party applications to other companies. They would benefit from the success of these companies who would drive more transactions over their rails.

Each of these strategies poses risks. To evaluate them it is important to agree on three premises concerning how the payments world will evolve over the next decade regardless of what the networks do.

First, over time ***the distinction between the online and offline world in payments will vanish***. Fast internet connections will become ubiquitous. Almost every device we interact with will have a connection. Those devices will access web-based software which will provide an increasing array of services.

We have already seen this happening. When the commercial internet started only a few computers were connected. Over time virtually all computers had internet connectivity. Then mobile phones started being connected. It is widely expected that most phones in industrialized countries will be internet-enabled in not too long. The third screen—the television—is next. Other devices are in the queue. Automobile manufacturers, for example, are incorporating internet capabilities in their cars and providing software platforms for applications. Importantly, for our purposes, point-of-sale devices will become internet-enabled and rely on services provided by web-based software. That will include everything from card terminals to cash registers.

This means that the neat distinction of today between online transactions that you do with your computer (or increasingly mobile phone) over the internet, and offline transactions where you walk into a brick-and-mortar store without touching the internet, will disappear. The internet will become like the electric power grid—something that almost every device is plugged into. The corollary is that it will be meaningless to talk about internet-payment providers, such as PayPal, as if they were an interesting group of foreigners. The alternative payment providers will be at the physical point of sale because the physical point of sale will be connected to the web.

Second, ***most devices even at the point of sale will rely on software, probably residing in the cloud, that provide valuable services***. At the moment most point-of-sale devices rely on software that either resides on the device or on servers that are accessed over a corporate intranet. Devices usually rely on a dedicated software solution. Once devices can obtain easy access to the internet it will prove more efficient to move software to the internet as well. Device makers may do this on their own to obtain greater control over software—including their choice of vendor and the ability to provide easy updates over the lifecycle of their equipment. Merchants may demand this as well because cloud-based software will provide them greater flexibility.

Innovation for the point-of-sale will move out to the cloud or will at least become untethered with the particular point-of-sale device. Internet-connected computers made it extremely easy for software developers to distribute products to businesses and people via download. That of course led to an increasingly vibrant application community. Internet-connected point of sale devices will do the same. The mere availability of the direct link to the point-of-sale devices will ignite a developer community.

Third, ***there's going to be limited room for software payments platforms.*** Several software payments platforms will succeed in becoming the go-to places for developers who want to incorporate payments and related functionality into their products and services. But it is doubtful that there's going to be room for more than two or three.

Software platforms are driven by indirect network effects. The more applications that are written for them the more customers (businesses and consumers in this case) will want to use them; and the more customers there are for the applications for a platform the more other developers will want to write more applications. The main reason this virtuous circle doesn't lead to a monopoly platform is that different platforms can specialize in providing different services—product differentiation can offset indirect network effects. The other limiting factor is that software developers will port their applications to a few platforms but not many. They will go for the two or three platforms that have the largest reach assuming they are technically on par.

For payments, it is doubtful that the four card networks plus several alternative payments providers plus others will all succeed in operating software payments platforms. This point has, as we will see, has huge implications for the risks faced by the networks in choosing a strategy.

With these premises in place it is easy to see the risks and opportunities that the card networks face during the new age of invisible engines.

Let's start with the stick-to-your knitting strategy of focusing on just running the rails. That would involve future value-added services moving to third-parties. The software payments platforms would support ecosystems of application developers that would incorporate payments functionality into their applications. (Just to be clear, these applications could be software programs that support substantial businesses. That's the vision for Square for example.) The applications businesses would earn revenue from their customers and the software platforms would earn license fees (and perhaps revenue cuts) from the applications business.

The card networks could charge for access to their rails. They would benefit from more applications generating more transactions over their networks. The risk is two-fold. On the one hand the software platforms and applications suck up a lot of the value that's available from using the rails. The card rails could try to get a piece of the action but that brings us to the other hand: the software platforms, and perhaps other players, could try to disintermediate the rails altogether. In fact, one would expect that PayPal will squeeze card fees down. They are already moving transactions over to cheaper-for-them ACH. Their ability to do so will eventually put downward pressure on transaction fees for the card networks (not to mention interchange fees if those survive the current regulatory and legal onslaughts). The card networks could try to be like the cable companies and squeeze a lot from everyone who tries to use their rails. That has worked, although it is getting ugly, in cable because there isn't much competition. It is unlikely to work in payments where there are multiple sets of payments rails and lot of opportunities (see PayPal above) to shift transactions in response to cost differences.

The card networks could also take the middle way. They could embrace the idea of creating applications that use their network services but do that themselves, but more aggressively than they have in the past and with a clear eye on what's being built on the open platforms, or enlist business partners to that with them. This has several advantages. For one, it allows them to keep tight control over what's built on their networks, make sure it is consistent with their overall business objectives, tightly integrate the applications, and prevent applications that could degrade the performance or reputation of their networks. For another, it allows them to make money from the applications directly rather than limiting themselves to access and transactions fees. They will own customer relationships and data as well. The chief disadvantage of the closed-platform approach is that it does not generate the sorts of indirect network effects that have powered the growth of the iPhone and seem to be working well so far for PayPal. The networks would have to hope that they make up in quality of apps what they lose in quantities.

Finally, the card networks could adopt an invisible engine strategy. That would generally involve creating a software layer above their networks, developing features for that layer that use their rails and provide other services as well, and making these features available through APIs. Following the PayPal X approach they would need to evangelize their platforms and provide developers with the tools and encouragement to create successful applications. There is a huge upside potential here. They could, as I've described throughout this series, become the centre of a ecosystem of applications that would drive value back to the center and in doing create barriers for rivals. See Windows.

Like most things in life that have major payoffs, there are at least two related and significant risks: First, it is highly unlikely that all four of the card networks could succeed in establishing successful software payments platforms. One could easily imagine that only 2-3 platforms emerge and that at least one of those is one of the alternative payments providers. That would mean that if all four networks tried 2-3 of them would probably fail. Second, there can be a wide gulf between aspiration and ability as every little boy who wants to play pro-ball soon finds out. The card networks have much experience developing sophisticated software systems and communications platforms. So it isn't like a car company wanting to get into the software platform business. But the card networks have evolved slowly from mainframe based companies and don't, at least obviously, have demonstrated skills in doing internet-related development work. They may lack the genetic matter to go up against PayPal or other companies that were born in the internet world. Thus, the card networks could make significant investments of financial and reputational capital in trying to become one of the software payments platforms and fail in succeeding.

The rise of invisible engines in payments therefore presents significant opportunities and risks for the traditional card networks. One or more of them could distance themselves from their rivals by becoming one of the surviving software payments platforms with a robust set of applications driving traffic. Those that don't develop a successful platform could find themselves as service providers to much more interesting and robust businesses. It is beginning to look like this is the way the mobile communications business could evolve—with the iPhone, the Android, and perhaps another software development platform driving applications and the carriers settling for charging for carrying traffic.

The traditional card networks are clearly trying to navigate what to do. Visa's acquisition of CyberSource for $2 billion may be an important step in its developing an open software platform. CyberSource is the e-commerce engine for about 300,000 merchants. It mainly helps these firms with key back office tasks—risk and fraud most importantly, call centers, and then processing. Combined with other technology it could help Visa challenge PayPal X. American Express's acquisition of Revolution Money for $300 million is another example. Revolution Money has a payment platform based on internet technology. It could also provide some useful building blocks for a larger efforts.

The decisions these networks make on how to proceed, and how to meet the Pay Pal X challenge, will determine the course of payments over the next decade.

IX. MOVING INNOVATION TO THE EDGE

Innovation has happened slowly in the payments industry. The experience with the introduction of contactless at the point-of-sale in the United States illustrates why. Merchants had to install new equipment to accept contactless cards. Many haven't wanted to do that without evidence that consumers want to use the card. But there hasn't been a groundswell of demand from consumers because there are few places to use the cards or really compelling reasons to use it. It is hard to ignite innovations in payments because it is costly to change current systems and because it is difficult to solve the chicken-and-egg problem of getting multiple stakeholders in the ecosystem on board.

Cloud-based software platforms may change that very quickly. Innovation will take place through the creation of software-enabled businesses that access the payments rails through software development platforms that live in the cloud. These innovations will be pushed through the broader payments ecosystem through the internet.

Whether Square succeeds or not as a business it points the way to how this will happen. It is based on a software application that relies on two software development platforms—IP Commerce's payments software development platform and Apple's iPhone software development platform. It connects to merchants through an internet-connected device—the iPhone, iPad, or other smart mobile technology. It connects to processors through an internet connection that takes it to IP Commerce's software platform in the cloud. And it connects to consumers through email that traverses the internet.

Rentalic—the winner of PayPal X's first application competition—may also succeed or fail but again points the way. An entrepreneur comes up with an idea for bringing buyers and sellers together (in this case renters and rentees) and needs a way to deal with complex payments (in this case handling the security deposit). She goes to a payments software platform in the cloud for these services. PayPal X enables Rentalic to connect buyers and sellers and ultimately to multiple payment rails.

Rentalic and Square demonstrate why the pace of innovation is likely to increase rapidly in payments and highlight the opportunities. Innovation will pick up because the cost of getting new ideas that require payments solutions is declining rapidly as a result of these software platforms. The opportunities are enormous—one only has to look at what Apple has accomplished with its Applications Store and glance at how quickly PayPal X has attracted developers and new applications.

The "applications" also highlight the risks facing the traditional card networks if they fall behind. Both Rentalic and Square involve physical transactions—the bread and butter of the card networks. Rentalic is mainly about renting physical items where the buyer and seller will actually meet in person probably. Square is about transactions at physical points of sale. That's the future. One or more of the card networks may figure out how to create an open software development platform and get a virtuous circle going around an "applications store". Some card networks may find themselves on the sidelines and watch a huge amount of value being sucked towards the cloud.

Invisible engines will drive innovation and transform the payments industry. Some incumbents will win, and some will lose, as this disruption unfolds over the next decade. But there's no doubt in my mind the entrepreneurs, consumers, retailers, and other stakeholders in the payments ecosystem will benefit tremendously

WHERE CHAPTERS WERE ORIGINALLY PUBLISHED

Chapter 1: "The Industrial Organization of Markets with Two-Sided Platforms," *Competition Policy International*, 3 (1) Spring 2007, with Richard Schmalensee.

Chapter 2: "Some Empirical Aspects of Multi-Sided Platform Industries" *Review of Network Economics*, Vol. 2, Issue 3, September 2003.

Chapter 3: "How Catalysts Ignite: The Economics of Platform-Based Startups," *Platforms, Markets and Innovation*, A. Gawer, ed., Cheltenham, UK and Northampton, MA, US: Edward Elgar, 2009.

Chapter 4: "The Antitrust Economics of Multi-Sided Platforms," *Yale Journal on Regulation*, Vol. 20, Issue 2, Summer 2003.

Chapter 5: "Two-Sided Market Definition," ABA Section of Antitrust Law, *Market Definition in Antitrust: Theory and Case Studies*, forthcoming 2012.

Chapter 6: "The Analysis of Mergers that Involve Multi-Sided Platforms", *Journal of Competition Law and Economics* 2008:4, pp. 663-695, with Michael D. Noel.

Chapter 7: "The Economics of the Online Advertising Industry," *Review of Network* Economics, Vol. 7, No. 3, September 2008.

Chapter 8: "The Online Advertising Industry: Economics, Evolution, and Privacy," *Journal of Economic Perspectives*, Summer 2009.

Chapter 9: "Web Economy, Two-Sided Markets, and Competition," *Concurrences,* May 2010

Chapter 10: "Antitrust Issues Raised by Global Internet Economy," *Northwestern University Law Review*, Vol. 102, No. 4, p. 285, Colloquy Essay, 2008.

Chapter 11: "More Than Money," *Paying with Plastic* (Massachusetts: MIT Press, first edition 1999, second edition 2005; third edition 2012), with R. Schmalensee.

Chapter 12: "Interchange Fees and Their Regulation," *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?*, Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 2005

Chapter 13: "The Effect of Regulatory Intervention in Two-Sided Markets," *Review of Network Economics*, Vol. 4, Issue 4, December 2005, with Howard H. Chang and Daniel D. Garcia Swartz.

Chapter 14: Invisible Engines: Introduction to the Chinese Edition *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries*, (Massachusetts: MIT Press, 2010) with A Hagiu and R. Schmalensee.

Chapter 15: "Invisible Engines: Both Sides Now," *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries*, (Massachusetts: MIT Press, 2006) with A Hagiu and R. Schmalensee.

Chapter 16: "What's Next in Payments: Invisible Engines," PYMNTS.com, October 24, 2009

BIBLIOGRAPHY

Ackerberg, Daniel A., and Gautam Gowrisankaran. 2003. "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," mimeo, University of Arizona, October.

Ahlborn, Christian, Howard H. Chang, and David S. Evans. 2001. "The Problem of Interchange Fee Analysis: Case without a Cause?" *European Competition Law Review*, vol. 22, no. 8 (August), pp. 304-312.

American Express. 2005. *American Express Company Annual Report 2004*. February.

Anderson, Simon P., Jean J. Gabszewicz. 2005. "The Media and Advertising: A Tale of Two-Sided Markets." *CEPR Discussion Paper No. 5223*.

ANZ (2003) "Annual Report."

Arango, Tim. 2008. "Cable Firms Join Forces to Attract Focused Ads," *New York Times* (March10).

Arcavision. "Exchange Volume Summary-1/1/2008-12/31/2008." Accessed on April 2, 2009.

Argentesi, E. and L. Filistrucchi (2005). "Estimating Market Power in a Two-Sided Market: The Case of Newspapers," Journal of Applied Econometrics, *forthcoming*.

Argentesi, E. and M. Ivaldi (2005). "Market Definition in Printed Media Industry: Theory and Practice," CEPR Discussion Paper No. 5096, Available at SSRN: http://ssrn.com/abstract=779107.

Armstrong, M. and J. Wright (2007). "Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts," Economic Theory, Vol. 32, No. 2, pp. 353-380.

Armstrong, Marcia (1991) "Retail Response to Trade Promotion: An Incremental Analysis of Forward Buying and Retail Promotion," unpublished doctoral dissertation, School of Management, University of Texas at Dallas

Armstrong, Mark (2002) "Competition in Two-Sided Markets," Nuffield College, Oxford, Working Paper.

Armstrong, Mark. 2004. "Competition in Two-Sided Markets," mimeo, University College London.

___________. 2005. "Two-Sided Markets: Economic Theory and Policy Implications." mimeo, University College, London, April.

Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *RAND Journal of Economics*, 37 (3), 668-691.

Armstrong, Mark (2006). "Competition in Two-Sided Markets," Rand Journal of Economics, Vol. 37, No. 3, pp. 668-691.

Armstrong, Mark, and Julian Wright. 2004. "Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts," mimeo, University College London, November.

Arthur, W. Brian (1989) "Competing Technologies, Increasing Returns, and Lock-In By Historical Events," *The Economic Journal*, 99: 116-131.

Atluru, Rajesh and Kevin Wasserstein (1998) "Palm Computing: The Pilot Organizer," Harvard Business School.

Australian Bureau of Statistics (2001) "2001 Census Basic Community Profile and Snapshot." Online: http://www.abs.gov.au/ausstats/abs@census.nsf/4079a1bbd2a04b80ca256b9d00208f92/7dd97c937216e32fca256bbe008371f0!OpenDocument

Australian Bureau of Statistics (2003) "Year Book Australia." Online:http://www.abs.gov.au/Ausstats/abs@.nsf/Lookup/361F400BCE3AB8ACCA256CAE00053FA4

Australian Bureau of Statistics (2005) "Table 20. Household Final Consumption Expenditure," March.

Australian Prudential Regulation Authority (2005) Monthly Banking Statistics, July. online: http://www.apra.gov.au/Statistics/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=9124

Bagwell, Kyle. 2002. "The Economic Analysis of Advertising." mimeo, Columbia University.

Baker, J., and T. Bresnahan (1985) "The Gains from Merger or Collusion in Product-Differentiated Industries," Journal of Industrial Economics, 33(4), pp. 427-444.

Balto, David A. 2000. "The Problem of Interchange Fees: Costs without Benefits?" *European Competition Law Review*, vol. 21, no 4 (April), pp. 215-224.

Baumol, William, John C. Panzar and Robert D. Willig (1982) *Contestable Markets and The Theory of Industry Structure*, Harcourt Brace Jovanovich, New York.

Baxter, William (1983) "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26: 541-588.

Baxter, William F. 1983. "Bank Exchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics*, vol. 26, no. 3 (October), pp. 541-588.

Bayot, Jennifer. 2003. "Credit Card Rewards Turn From Nouns to Verbs," *New York Times*, September 21.

Becker, David (2000) "Revenue from Game Consoles will Plunge, Report Predicts," *CNET News.com*, July 31, http://techrepublic-cnet.com.com/2100-1040-43841.html?legacy=cnet.

Becker, David (2002) "Xbox Drags on Microsoft Profit," *CNET News.com,* Jan. 18, http://news.com.com/2100-1040-818798.html.

Bergin, Richard (1998) "Interview with Yahoo! Senior Vice President Tim Brady," Harvard Business School, doctoral research.

Bergman, Mats A. 2005. "Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs." Mimeo, Swedish Competition Authority, January.

Berndt, Ernst, Mohammad Arzaghi, James C. Davis, Alvin J. Silk. 2008. "Economic Factors Underlying the Unbundling of Advertising Agency Services." *NBER Working Paper No. W14345*.

Besley, Timothy and Harvey Rosen (1999) "Sales Taxes and Prices: An Empirical Analysis," *National Tax Journal*, 52(2): 157-178.

BIGresearch, Simultaneous Media Usage Study (SIMM12), Media Influence to Purchase & New Media Usage –Voters, June 2008.

Blodget, Henry. 2008. "Newspapers' Fab 2007: $11 Billion of Market Value Vaporized." Silicon Alley Insider. (January 2), at <http://www.alleyinsider.com/2008/1/newspapers_11_billion_of_market_value_vaporized_in_2007>.

Bolt, Wilko, and Alexander F. Tieman. 2004a. "Skewed Pricing in Two-Sided Markets: An IO Approach," Working Paper No. 13, De Nederlandsche Bank, Amsterdam, October.

___________. 2004b. "A Note on Social Welfare and Cost Recovery in Two-Sided Markets," Working Paper No. 24, De Nederlandsche Bank, Amsterdam, December.

Borenstein, Severin, C. Cameron, and R. Gilbert (1997) "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?," *Quarterly Journal of Economics* 112 (February).

Brammall, Bruce (2002). "Watchdog Fears ANZ Card Hikes," *Herald Sun*, November 5.

Brammall, Bruce (2003) "Bank May Ditch Reward Scheme," *Herald Sun*, February 20.

Bulow, Jeremy and Paul Pfleiderer (1983) "A Note on the Effect of Cost Changes on Prices," *Journal of Political Economy* 91(1): 182-185.

Bureau of Transportation Statistics (2000) "Airport Activity Statistics of Certificated Air Carriers: Summary Tables 2000," Table 1, online: http://www.bts.gov/publications/airport_activity_statistics_of_certificated_air_carri ers/2000/tables/table01.html.

Bureau of Transportation Statistics. 2000. "Airport Activity Statistics of Certificated Air Carriers: Summary Tables 2000," Table 1, online: http://www.bts.gov/publications/airport_activity_statistics_of_certificated_air_carriers /2000/tables/table01.html.

Caillaud, Bernard and Bruno Jullien (2001) "Chicken & Egg: Competing Matchmakers," Center For Economic Policy Research, Working Paper No. 2885.

Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken and Egg: Competition among Intermediation Service Providers," *Rand Journal of Economics*, vol. 34, no. 2 (Summer), pp. 521-552.

Cannex Australia (2004) Data on merchant service fees. August/September.

Carlton, Dennis (2007). "Market Definition: Use and Abuse," Competition Policy International, Vol. 3, No 1.

Carlton, Dennis W. (1986) "The Rigidity of Prices," *American Economic Review,* 76(4): 637-58

Carlton, Dennis W., and Alan S. Frankel. 1995a. "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, vol. 63, no. 2 (Winter), pp. 643-668.

___________, and ___________. 1995b. "The Antitrust Economics of Credit Card Networks: Reply to Evans and Schmalensee," *Antitrust Law Journal*, vol. 63, no. 3 (Spring), pp. 903-915.

Chakravorti, Sujit (2003) "Theory of Credit Card Networks: A Survey of the Literature," *Review of Network Economics*, 2(2): 50-68.

Chakravorti, Sujit. 2003. "Theory of Credit Card Networks: A Survey of the Literature," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 50-68.

Chakravorti, Sujit, and Alpha Shah. 2003. "Underlying Incentives in Credit Card Networks," *Antitrust Bulletin*, vol. 48, no. 1 (Spring), pp. 53-75.

Chakravorti, Sujit, and Roberto Roson. 2004. "Platform Competition in Two-Sided Markets: The Case of Payment Networks," Federal Reserve Bank of Chicago, Working Paper 2004-09, July.

Chakravorti, Sujit, and Ted To. 2003. "A Theory of Credit Cards," mimeo, Federal Reserve Bank of Chicago, July.

Chakravorti, Sujit, and William R. Emmons. 2003. "Who Pays for Credit Cards?" *Journal of Consumer Affairs*, vol. 37, no. 2 (Winter), pp. 208-230.

Chang, Howard H., and David S. Evans. 2000. "The Competitive Effects of the Collective Setting of Interchange by Payment Card Systems," *Antitrust Bulletin*, vol. 45, no. 3 (Fall), pp. 641-677.

Chang, Howard H., David S. Evans, and Daniel Garcia Swartz. 2005. "An Assessment of the Reserve Bank of Australia's Interchange Fee Reforms," Working Paper, May.

Chevalier, Michel and R. Curhan (1976) "Retail Promotions as a Function of Trade Promotions: A Descriptive Analysis," *Sloan Management Review* 18(3): 19-32.

Chiaravalle, Bill and Barbara Schenck. 2006. *Branding for Dummies*. John Wiley & Sons.

Cisneros, Oscar S (2000) "Ebay Accused of Monopolization," *Wired News*, July 31, http://www.wired.com/news/print/0,1294,37871,00.html.

ComScore Press Release. 2008. "U.S. Search Engines Rankings." (November 26), at <http://www.marketwatch.com/news/story/comScore-Releases-October-2008-US/story.aspx?guid={FE165F0B-907D-4869-BDC9-379F061AEB6E}>.

Cornell, Andrew. 2004. "Deal Highlights Change in Card Market," *Australian Financial Review*, July 12.

Cornell, Andrew. 2005. "American Express Aims for Expansion," *Australian Financial Review*, April 18.

Cotterill, Ronald (1998) "Estimation of Cost Pass Through to Michigan Consumers in the ADM Price Fixing Case," Food Marketing Policy Center, University of Connecticut, Research Report No. 39.

Cotterill, Ronald (2000) "Dynamic Explanations of Industry Structure and Performance," presented at the USDA Conference "The American Consumer and the Changing Structure of the Food System," May 3-5, 2000.

Cruickshank, Don. 2000. "Annex D3: Interchange Fees for Credit and Debit Cards," *UK Banking: A Report to the Chancellor of the Exchequer*, UK: HM Treasury (March), pp. 247-272.

De Grauwe, Paul, and Laura Rinaldi. 2002. "A Model of the Card Payment System and The Interchange Fee," in Clemens Esser and Michael H. Stierle, eds., *Current Issues in Competition Theory and Policy*, Berlin: VWF.

DOJ and FTC (1997). "U.S. Horizontal Merger Guidelines." Available at: http://www.usdoj.gov/atr/public/guidelines/horiz_book/hmg1.html.

Duboff, Robert S. 2007. *ROI for Marketing: Balancing Accountability with Long-Term Needs*. New York: Association of National Advertisers.

Dubois, P., A. Hernandez -Perez and M. Ivaldi (2007). "The Market of Academic Journals: Empirical Evidence from Data on French Libraries," Journal of the European Economic Association, Vol. 5, No. 2, April-May 2007, pp. 390-399

Dye, Jessica. 2009. "Consumer Privacy Advocates Seek Search Engine Solution." EContent (March Issue), at <http://www.econtentmag.com/Articles/News/News-Feature/Consumer-Privacy-Advocates—Seek-Search-Engine-Solution-52679.htm>.

Economics of Payment Card Associations," *Rand Journal of Economics*, 33: 549-570.

Edelman B., Michael Ostrovsky, and Michael Schwarz. 2007. "Internet Advertising and the Generalized Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review*, 97(1), 242-259.

Elliott, Stuart. 2006. "At the Four A's, a Focus on Getting Stronger." *New York Times* (April 3), at <http://www.nytimes.com/2006/04/03/business/03adcol.html?pagewanted=print>.

Enders, Walter (2004) *Applied Econometric Time Series*. New York: John Wiley and Sons.

Euromonitor International (2005a) "Department Stores in Australia," October.

Euromonitor International. 2005a. "Department Stores in Australia," October.

Euromonitor International (2005b) "Grocery Stores, Food Retailers and Supermarkets in Australia," October.

Euromonitor International. 2005b. "Grocery Stores, Food Retailers and Supermarkets in Australia," October.

Euromonitor International (2005c) "Superstores and Warehouse Clubs in Australia," October.

Euromonitor International. 2005c. "Superstores and Warehouse Clubs in Australia," October.

European Commission. 2002. "Commission Decision of 24 July 2002 Relating to a Proceeding Under Article 81 of the EC Treaty and Article 53 of the EEA Agreement, Case COMP/29.373—Visa International (2002/914/EC).

___________. 2003. "Statement on Mastercard." Midday Express Press Release, October 30.

European Commission. 2008.*Opinion on Data Protection Issues Related to Search Engines*, April.

Evans, D. (2003a). "The Antitrust Economics of Two-Sided Markets," Yale Journal of Regulation, Vol. 20, No. 2, pp. 325-382.

Evans, D. (2003b). "Some Empirical Aspects of Multi-Sided Platform Industries," Review of Network Economics, Vol. 2, No. 3, pp. 191-209.

Evans, D., A. Hagiu and R. Schmalensee (2006). *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries.* MIT Press, 2006.

Evans, D. and R. Schmalensee (2005b). "The Economics of Interchange Fees and Their Regulation: An Overview." Proceedings – Payment Systems Research Conferences, Federal Reserve Bank of Kansas City, May 2005, pp. 73-120.

Evans D. and R. Schmalensee (2007a). "The Industrial Organization of Two-Sided Platforms," Competition Policy International, Vol. 3, No. 1.

Evans, D. and R. Schmalensee (2007b). Catalyst Code: *The Strategies behind the World's Most Dynamic Companies.* Harvard Business School Press, 2007.

Evans, David and Richard Schmalensee (2005a). *Paying with Plastic: The Digital Revolution in Buying and Borrowing,* 2d Ed. MIT Press, Cambridge, MA.

Evans, David and Richard Schmalensee. 2005. *Paying with Plastic, 2nd Edition: The Digital Revolution in Buying and Borrowing.* Cambridge: The MIT Press.

Evans, David and Richard Schmalensee. 2009. "Platform Failure." Available at SSRN, http://ssrn.com/abstract=1353502.

Evans, David S. 2003a. "Some Empirical Aspects of Multi-Sided Platform Industries," *Review of Network Economics*, vol. 2, issue 3 (September), 191-209.

___________. 2003b. "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation*, vol. 20, no. 2 (Summer), pp. 325-381.

___________. 2003b. "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation*, vol. 20, no. 2 (Summer), pp. 325-381.

Evans, David S. (2003) "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation* (forthcoming).

Evans, David S. 2008. "The Economics of the Online Advertising Industry," *Review of Network Economics*, 7(3), pp 359 – 391.

Evans, David S. 2009. "How Catalysts Ignite: The Economics of Platform-Based Start-Ups." Platforms, Markets and Innovation, A. Gawer, ed., Cheltenham, UK and Northampton, MA, US: Edward Elgar, 2009. Available at SSRN: http://ssrn.com/abstract=1279631.

Evans, David S., Andrei Hagiu, and Richard Schmalensee. 2004. "A Survey of the Economic Role of Software Platforms in Computer-Based Industries," CESifo Working Paper No. 1314, October.

Evans, David S., and Richard Schmalensee. 1995. "Economic Aspects of Payment Card Systems and Antitrust Policy toward Joint Ventures," *Antitrust Law Journal*, vol. 63, no. 3 (Spring), pp. 861-901.

Evans, David S. and Richard Schmalensee (1999) *Paying with Plastic: The Digital Revolution in Buying and Borrowing*, The MIT Press, Cambridge, MA.

Evans, David S., and Richard Schmalensee (2005a) *Paying with Plastic: The Digital Revolution in Buying and Borrowing 2d Ed*. Cambridge, MA: MIT Press

Evans, David S., and Richard Schmalensee (2005b) "The Economics of Interchange Fees and Their Regulation: An Overview," paper presented at a Conference *Interchange* Fees in Credit and Debit Card Industries: What Role for Public Authorities? Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6.

Evans, David S., and Richard Schmalensee (2005c) "The Industrial Organization of Markets with Two-Sided Platforms," *NBER Working Paper Series*, No. 11603, September.

Fahey, Rob (2002) "MS to Lose J525M on Xbox This Year," *GameIndustry.biz*, June 26, http://www.gamesindustry.biz/content_page.php?section_name=pub&aid=210.

Farrell, J and C. Shapiro (1990). "Horizontal Mergers: An Equilibrium Analysis," The American Economic Review, Vol. 80, No. 1, pp. 107-126.

Federal Trade Commission Hearing. *Behavioral Advertising: Tracking, Targeting, and Technology*. November 2006.

Federal Trade Commission Release, *Online Behavioral Advertising: Moving the Discussion Forward to Possible Self-Regulatory Principles*. December 2007.

Feldman, R. (1994) "The Welfare Economics of a Health Plan Merger," Journal of Regulatory Economics, 6(1), pp. 67-86.

Feldstein, Janet and Christopher Flanagan (2001) "Handspring – Partnerships," Harvard Business School.

Ferguson, James M. (1983) "Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition," *Journal of Law and Economics*, 26: 635-654.

Festa, Paul (1996) "Looking for Love Online," *CNET News.com*, December 17, 1996 http://news.com.com/2100-1023-255523.html?tag=rn.

Fisher, James (1997) "An Ad Blitz for the 21st Century," Sprint Press Release, September 24, http://www3.sprint.com/PR/CDA/PR_CDA_Press_Releases_Detail/0,3245,,00.html?ID=1294.

Food Marketing Institute. 1998. *A Retailer's Guide to Electronic Payment System Costs,* Washington, DC: Food Marketing Institute.

Frankel, Alan S., 1998. "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, vol. 66, no. 2 (Winter), pp. 313-361.

Frew, J. and G. Jud (1987) "Who Pays the Real Estate Broker's Commission?" *Research in Law and Economics: The Economics of Urban Property Rights*, 10: 177-188.

Friedman, Jon and John Meehan (1992) *House of Cards: Inside the Troubled Empire of American Express*, Putnam, New York.

Gabszewicz, Jean J. and Xavier Y. Wauthy. 2004. "Two-Sided Markets and Price Competition with Multi-homing," mimeo, CORE, Catholic University of Louvain, Belgium.

Gans, Joshua S., and Stephen P. King. 2003a. "Approaches to Regulating Interchange Fees in Payment Systems," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 125-145.

__________, and __________. 2003b. "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis and Policy*, vol. 3, no. 1, online: http://www.bepress.com/bejeap/topics/vol3/iss1/art1/.

__________, and __________. 2003c. "A Theoretical Analysis of Credit Card Reform in Australia," *Economic Record*, v. 79, issue 247 (December), pp. 462-472.

Garcia Swartz, Daniel D., Robert W. Hahn, and Anne Layne-Farrar. 2004. "The Economics of a Cashless Society: An Analysis of the Costs and Benefits of Payment Instruments," AEI-Brookings Joint Center for Regulatory Studies, Related Publication 04-24, September.

Gawer, Annabelle and Michael A. Cusumano (2002) *Platform Leadership*, Harvard Business School Press, Boston, MA.

George, Lisa and Joel Waldfogel (2000) "Who Benefits Whom in Daily Newspaper Markets?" NBER Working Paper #7944.

Goettler, Ronald. (1999) "Advertising Rates, Audience Composition, and Competition in the Network Television Industry," Carnegie Mellon University GSIA Working Paper #1999-E28.

Goldberger, Arthur S. (1991) *A Course in Econometrics*, Cambridge, Massachusetts: Harvard University Press.

Goldberg, Pinelopi and Michael Knetter (1997) "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature* 35(3): 1243-1272.

"Google's growth raises privacy concerns." 2008. *CNN.com*. November 4.

"GoTo to Overture to YSM-Timeline," 2005. September 21. bg Theory website, at <http://www.bgtheory.com/blog/goto-to-overture-to-ysm-timeline/>.

Graeme, James (2003) "Loyalty is No Longer Paying Off," *Sunday Telegraph*, May 4.

Greene, William H (1993) *Econometric Analysis.* New York: Macmillan, 1993.

Grossman, Peter (1987) *American Express: The Unofficial History of the People Who Built the Great Empire*, Crown Publishers, New York.

Guerrera, Francesco (2002) "Brussels Set to Let Visa off the Hook," *Financial Times*, May 13.

Guthrie, Graeme, and Julian Wright. 2003. "Competing Payment Schemes," mimeo, National University of Singapore, September.

Hacker, Scot (2001) "He Who Controls the Bootloader," *Byte*, August, http://www.byte.com/documents/s=1115/byt20010824s0001/0827_hacker.html.

Hagel, John and Marc Singer. 1999. *Net Worth*. Cambridge: Harvard Business School Press.

Hagiu, Andrei. 2004a. "Two-Sided Platforms: Pricing and Social Efficiency," RIETI Discussion Paper Series 04-E-035, Research Institute for the Economy, Trade and Industry, Tokyo, December.

___________. 2004b. "Two-Sided Platforms: Pricing and Social Efficiency – Extensions," RIETI Discussion Paper Series 04-E-036, Research Institute for the Economy, Trade and Industry, Tokyo, December.

Hallerman, David. 2008. "US Online Advertising: Resilient in a Rough Economy: Summary." March, available at eMarketer.com, <http://www.emarketer.com/Reports/All/Emarketer_2000488.aspx>.

Hamilton, James (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hammer, Alexander R. (1962) "American Express Raises Profit: Credit Card Unit in the Black," *The New York Times*, April 25.

Hanna, Jim (2002) "Fed: Consumers to Save $500m-Plus Under Credit Card Shake-Up," *AAP Newsfeed*, August 27.

Hansen, Bruce (2001) "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives*, 15(4).

Harris, B. (2003). "Recent Observations about Critical Loss Analysis." Available at: http://www.usdoj.gov/atr/public/workshops/docs/202599.htm

Harris, B. and C. Veljanovski (2003). "Critical Loss: Its Growing Use in Competition Law," European Competition Law Review, Vol. 5, pp. 213-218.

Harris, B. and J.J. Simons (1989). "Focusing Market Definition: How Much Substitution is Necessary?," Research in Law and Economics, Vol. 207, pp. 207-226.

Harris, Larry. 2002. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press.

Hayashi, Fumiko. 2004. "A Puzzle of Payment Card Pricing: Why Are Merchants Still Accepting Card Payments?" mimeo, Federal Reserve Bank of Kansas City, December.

Hayashi, Fumiko, Richard Sullivan, and Stuart E. Weiner. 2003. *A Guide to the ATM and Debit Card Industry*, Federal Reserve Bank of Kansas City.

Helft, Miguel. 2008."Google's New Tool Is Meant for Marketers." *The New York Times*. (August 6), at http://www.nytimes.com/2008/08/06/business/media/06adco.html.

"History of Pay Per Click Advertising and Yahoo! Sponsored Search," 2007. November 1. Semvironment.com website at <http://www.semvironment.com/the-history-of-pay-per-click-advertising-and-yahoo-sponsored-search>.

Horan, Matthew (2003) "Banks Up Card Fees Yet Again," *Sunday Telegraph*, September 21.

House Of Representatives Standing Committee On Economic Economics, Finance And Public Administration (2005) Official Committee Hansard. August 12, online: http://www.aph.gov.au/hansard/reps/commttee/R8516.pdf

HSN Consultants. 1987. *The Nilson Report*, no. 406 (June).

__________. 1991. *The Nilson Report*, no. 510 (October).

__________. 1992. *The Nilson Report*, no. 522 (April).

__________. 2001a. *The Nilson Report*, no. 737 (April).

__________. 2001b. *The Nilson Report*, no. 738 (April).

HSN Consultants (2003) *The Nilson Report*, no. 783, March.

Huber, Peter J. (19640 "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35(1): 73-101.

Humphrey, David, Magnus Willesson, Ted Lindblom, and Göran Bergendahl. 2003. "What Does it Cost to Make a Payment?" *Review of Network Economics*, vol. 2, no. 2 (June), pp. 159-174.

Hunt, Robert M. 2003. "An Introduction to the Economics of Payment Card Networks," *Review of Network Economics*, vol. 2, no. 2 (June), 80-96.

IMA Market Development AB. 2000. "Study Regarding the Effects of the Abolition of the Non-discrimination Rule in Sweden: Final Report, Results and Conclusions," mimeo, Lerum, Sweden, February.

Interactive Advertising Bureau. 2000-2007. Press Releases.

ITM Research. 2000. "The Abolition of the No-discrimination Rule." mimeo, Amsterdam, March.

Ivaldi, M. and F. Verboven (2005). "Quantifying the Effects from Horizontal Mergers in European Competition Policy," International Journal of Industrial Organization, Vol. 23, pp. 669-691.

Jackson, William, III (1997) "Market Structure and the Speed of Price Adjustments: Evidence of Non-monotonicity," *Review of Industrial Organization*, 12.

Jeon, Doh-Shin, Jean-Jacques Laffont, and Jean Tirole (2001) "On the Receiver Pays Principle," mimeo.

Johnston, Jack (1983) *Econometric Methods.* New York: McGraw-Hill.

Jullien, Bruno (2001) "Competing in Network Industries: Divide and Conquer," Institut D'Economie Industrielle, Working Paper No. 9.

Jullien, Bruno. 2004. "Two-Sided Markets and Electronic Intermediaries," CESifo Working Paper No. 1345, November.

Kaiser, U. and J. Wright (2006). "Price Structure in Two-Sided Markets: Evidence from the Magazine Industry," International Journal of Industrial Organization, Vol. 24, Issue 1, pp. 1-28.

Karrenbrok, Jeffrey D. (1991) "The Behavior of Retail Gasoline Prices: Symmetric or Not?," *Federal Reserve Bank of St. Louis Review*, 73.

Katz, M. and C. Shapiro (2003). "Critical Loss: Let's Tell the Whole Story," Antitrust Magazine, Vol. 17, No. 2, pp. 49-56.

Katz, Michael L. (2001) "Network Effects, Interchange Fees, and No-Surcharge Rules in the Australian Credit and Charge Card Industry," Commissioned Report in *Reform of Credit Card Schemes in Australia*, Reserve Bank of Australia, August.

Katz, Michael L. 2001. *Reform of Credit Card Schemes in Australia II, Commissioned Report: Network Effects, Interchange Fees, and No-Surcharge Rules in the Australian Credit and Charge Card Industry*, Reserve Bank of Australia, August.

Katz, Michael L., and Harvey S. Rosen (1998) *Microeconomics.* Boston, Massachusetts: Irwin/McGraw-Hill.

Kaye, Barabara K and Norman Medoff. 2001. *Just A Click Away: Advertising on the Internet*. Massachusetts: Allyn and Bacon.

Klein, B., Andres V. Lerner, Kevin M. Murphy and Lacey L. Plache (2006). "Competition in Two-Sided Markets: The Antitrust Economics of Payment Card Interchange Fees," Antitrust Law Journal, 73(3), pp. 571-626.

Konkurrencestyrelsen. 2005. "Konkurrencestyrelsens Undersøgelse af Dankortgebyret," ISA/1107-8900-003/bjs/fab/lkf, Copenhagen, February.

Lafayette, Jon. 2008. "Nielsen Gains Access to Data From Charter's Set-Top Boxes," TVWeek.com. (March 12), at <http://www.tvweek.com/news/2008/03/nielsen_gains_access_to_data_f.php>.

Lerner, Abba (1934) "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, 1: 157-175.

Lerner, Josh (2002) "Did Microsoft Deter Software Innovation?" Working Paper, http://papers.ssrn.com/sol3/delivery.cfm/SSRN_ID269498_code010528670.pdf?abstractid=2 69498.

Lipsey, R.G., and Kelvin Lancaster. 1956. "The General Theory of the Second Best," *Review of Economic Studies*, vol. 24, no. 1 (October), pp. 11-32.

Littlechild, Stephen. 2003. "Reflections on Incentive Regulation," *Review of Network Economics*, vol. 2, no. 4 (December), pp. 289-315.

Lucking-Reilly, David. and Daniel F. Spulber (2001) "Business-to-Business Electronic Commerce," *Journal of Economic Perspectives*, 15: 55-68.

Macfarlane, I.J. 2005. "Gresham's Law of Payments," Talk to AIBF Industry Forum 2005, Sydney, 23 March.

MacKintosh, James (2001) "Watchdog Rules Credit Card Fees to Retailers Excessive," *Financial Times*, September 26.

Mandell, Lewis (1990) *The Credit Card Industry: A History*, Twayne Publishers, Boston, MA.

Manenti, Fabio M., and Ernesto Somma. 2003. "Plastic Clashes: Competition among Closed and Open Systems in the Credit Card Industry," mimeo, Università di Padova, October.

Marketing Sherpa Report. 2008. "Banner Ad Size and Click Rate: Bigger a Bit Better, But It is the Click that Counts," September 16. "The Medium Is The Massages." 2002. *Adweek*, July 29.

Maxwell, Miranda (2005) "Tough Leader Wanted for Australia's Telstra," *Reuters*, May 10, online: http://www.reuters.com/newsArticle.jhtml?type=reutersEdge&storyID=8443245.

Maxwell, Miranda. 2005. "Tough Leader Wanted for Australia's Telstra," *Reuters*, May 10, online: http://www.reuters.com/newsArticle.jhtml?type=reutersEdge&storyID=8443245.

McFarland, Lyndal. 2005. "Update: Australia's Virgin Blue Outlook Clouded by Fuel," *CNN Money*, May 18, online: http://money.cnn.com/services/tickerheadlines/for5/200505180450 DOWJONESDJONLINE000368_FORTUNE5.htm.

McKinnon, Michael (2001) "Credit Card Profit Under Fire," *Courier Mail*, April 27.

Megginson, William L., and Jeffrey N. Netter. 2001. "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature*, vol. 39, no. 2 (June), pp. 321-389.

Microsoft Advertising Industry News. 2008. "Spending on Behavioral Targeting Predicted to Increase," September 8. At <http://advertising.microsoft.com/news-events/behavioral-targeting>.

Minton, Stephen. 2007. *Worldwide Telecom Spending 2007-2011 Forecast: World Telecom Black Book*. IDC.

National Bancard Corporation v. Visa U.S.A., Inc. 1982. "Defendant's Trial Brief," Southern District of Florida, April 12.

___________. 1984. 596 F. Supp. 1231 (S.D. Fla.).

___________. 1986a. 779 F.2d 592 (11th Cir.).

___________. 1986b. 479 U.S. 923.

Network Economics Consulting Group (2005) "Early evidence of the impact of Reserve Bank of Australia regulation of open credit card schemes," May.

Nevo, A. (2000). "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," RAND Journal of Economics, 31(3), pp. 395-421.

Newmark, David, and Steven Sharpe (1992) "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposits," *Quarterly Journal of Economics*, 107.

Nocera, Joseph. (1994) *A Piece of the Action: How the Middle Class Joined the Money Class*, Simon & Schuster, New York.

O'Brien, D. and A.Wickelgren. (2003). "A Critical Analysis of Critical Loss," FTC Working Paper 254.

Office of Fair Trading. 2003a. "MasterCard Interchange Fees: Preliminary Conclusions," London, OFT634, February.

Office of Fair Trading. 2003b. "UK Payment Systems: An OFT Market Study of Clearing Systems and Review of Plastic Card Networks," London, OFT658, May.

O'Hara, Maureen. 1998. *Market Microstructure Theory*. Oxford: Blackwell Publishing Inc.

Pacelle, Mitchell, Jathon Sapsford and Kara Scannell (2003) "Visa to Settle With Retailers for $2 Billion," *The Wall Street Journal*, May 1.

Parker, Geoffrey and Marshall W. Van Alstyne (2002) "Unbundling the Presence of Network Externalities," Working Paper, http://www.idei.asso.fr/Commun/Conferences/Internet/Janvier2003/Papiers/VanAl.pdf.

Pashigian, B. Peter and Eric D. Gould (1998) "Internalizing Externalities: The Pricing of Space in Shopping Malls," *Journal of Law and Economics*, 41: 115-142.

Paul, Ryan. 2008. "EU calls phoul over ad company Phorm's invasive snooping," Ars Technica (August 15), available at http://arstechnica.com/old/content/2008/08/eu-calls-phoul-over-ad-company-phorms-invasive-snooping.ars.

Peltzman, Sam (2000) "Prices Rise Faster than they Fall," *Journal of Political Economy*, 108(3).

Perron, Pierre (1989) "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57(6): 1361-1401.

Poirier, Dale J. (1976) *The Econometrics of Structural Change*. Amsterdam: North Holland.

Poirier, Dale J. and Steven G. Garber (1974) "The Determinants of Aerospace Profit Rates 1951-1971," *Southern Economic Journal*, 41(2): 228-238.

Pope, Daniel. 1983. *Making of Modern Advertising.* New York, NY: Basic Books.

Poterba, James (1999) "Retail Price Reactions to Changes in State and Local Sales Taxes," National Tax Journal, 49(2).

Reserve Bank of Australia (2001) "Reform of Credit Card Schemes in Australia I: A Consultation Document," December.

Reserve Bank of Australia. 2001. "Reform of Credit Card Schemes in Australia I: A Consultation Document," December.

___________. 2002. "Reform of Credit Card Schemes in Australia IV: Final Reforms and Regulation Impact Statement," August.

___________. 2004. "Payment System Board Annual Report 2004."

___________. 2005a. "Reform of the EFTPOS and Visa Debit Systems in Australia: A Consultation Document," February.

___________. 2005b. "Additional Credit Card Statistics," May, online: http://www.rba.gov.au/PaymentsSystem/AustralianPaymentsSystem/ExcelFiles/RPS.xls.

Reserve Bank of Australia (2002) "Reform of Credit Card Schemes in Australia IV: Final Reforms and Regulation Impact Statement," August.

Reserve Bank of Australia (2004) "Payment System Board Annual Report 2004."

Reserve Bank of Australia (2005a) "Banking Fees in Australia."

Reserve Bank of Australia (2005b) "Table C01: Credit and Charge Card Statistics." online: http://www.rba.gov.au/Statistics/Bulletin/C01hist.xls.

Reserve Bank of Australia (2005c) "Additional Credit Card Statistics." Online: http://www.rba.gov.au/PaymentsSystem/PaymentsStatistics/ExcelFiles/RPS.xls

Reserve Bank of Australia (2005d) "Market Shares Of Credit And Charge Card Schemes." Online: http://www.rba.gov.au/Statistics/Bulletin/C02hist.xls.

Reserve Bank of Australia and Australian Competition and Consumer Commission (2000) "Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access," October.

Rochet, J-C. and J. Tirole (2003). "Platform Competition in Two-Sided Markets," Journal of the European Economic Association, 1:4, pp. 990-1029.

Rochet, J-C. and J. Tirole (2006). "Two-Sided Markets: A Progress Report," RAND Journal of Economics, Vol. 37, Issue. 3, pp. 645-667.

Rochet, Jean-Charles (2003) "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, 2(2): 97-124.

Rochet, Jean-Charles. 2003. "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 97-124.

Rochet, Jean-Charles (2003) "The Theory of Interchange Fees: A Synthesis of Recent Contributions," Working Paper.

Rochet, Jean-Charles (2005) "The Interchange Fee Mysteries—Discussion of 'Economic Rationale for Interchange Fees' by D. Evans and R. Schmalensee," presented at a Conference *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6.

Rochet, Jean-Charles and Jean Tirole (2002) "Cooperation Among Competitors: Some

Rochet, Jean-Charles, and Jean Tirole. 2002. "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, vol. 33, no. 4 (Winter), pp. 549-570.

__________, and __________. 2003a. "An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 69-79.

__________, and __________. 2003b. "Platform Competition in Two-Sided Markets," *Journal of the European Economics Association*, vol. 1, no 4 (June), pp. 990-1209.

__________, and __________. 2004a. "Defining Two-Sided Markets," mimeo, IDEI, Toulouse, France, January.

__________. and __________. 2004b. "Two-Sided Markets: An Overview." Mimeo, IDEI, Toulouse, France, March.

Rochet, Jean-Charles and Jean Tirole (2002) "Cooperation among Competitors: Some Economics of Payment Card Associations," *The RAND Journal of Economics*, 33(4): 1-22.

Rochet, Jean-Charles and Jean Tirole (2003) "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1(4): 990-1029.

Rochet, Jean-Charles and Jean. Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association*, 1(4), 990-1029.

Rochet, Jean-Charles and Jean Tirole (2003) "Platform Competition in Two-Sided Markets," *Journal of the European Economics Association* (forthcoming).

Rochet, Jean-Charles and Jean. Tirole. 2006. "Externalities and Regulation in Card Payment Systems." *Review of Network Economics* 5(1), 1-14.

Ronan, Courtney (1998) "Apartment Locaters: How Do They Make Their Money?" *Realty Times*, June 30, http://realtytimes.com/rtnews/rtcpages/19980630_aptlocator.htm.

Rysman, M. (2004). "Competition Between Networks: A Study of the Market for Yellow Pages," Review of Economic Studies, 71, pp. 483-512.

Rysman, M. (2007). "An Empirical Analysis of Payment Card Usage," Journal of Industrial Economics, 55, pp. 1-36.

Rysman, Marc (2002) "Competition Between Networks: A Study of the Market for Yellow Pages 1-2," Boston University Industrial Studies Project, Working Paper No. 104.

Rysman, Marc. 2004. "An Empirical Analysis of Payment Card Usage," mimeo, Boston University, January.

Scheffman, D. and J. Simons (2003). "The State of Critical Loss: Let's Make Sure We Understand the Whole Story," The Antitrust Source, Vol. 3, No. 2, pp. 1-9.

Scherreik, Susan (2002) "Is Your Broker Leaving You Out in the Cold?" *Business Week*, http://www.businessweek.com/magazine/content/02_07/b3770110.htm.

Schiff, Aaron (2003) "Open and Closed Systems of Two-Sided Networks," *Information Economics and Policy*, forthcoming.

Schiff, Aaron. 2003. "Open and Closed Systems of Two-Sided Networks," *Information Economics and Policy*, vol. 15, no. 4 (December), pp. 425-442.

Schmalensee, Richard (2002) "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50(2): 103-122.

Schmalensee, Richard. 2002. "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, vol. 50, no. 2 (June), pp. 103-122.

___________. 2003. "Interchange Fees: A Review of the Literature," *Payment Card Economics Review*, vol. 1, no. 1 (Winter), pp. 25-44.

Schmalensee, Richard (2003) "Interchange Fees: A Review of the Literature," *Payment Card Economics Review*, 1: 25-44.

Schumpeter, Joseph. 1942. *Capitalism, Socialism and Democracy*. New York: Harper and Brothers.

Schwartz, Marius, and Daniel R. Vincent. 2004. "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," mimeo, Georgetown University.

Sears, Jay. 2005. "The Evolution of Contextual Advertising," Upgrade Magazine. (April/May), at <http://www.contextweb.com/pdf/UpgradeMagazine.pdf>.

Shapiro, Carl (1996). "Mergers with Differentiated Products," Antitrust, 10, pp. 23-30.

Shapiro, Carl and Hal R. Varian (1999) *Information Rules*, Harvard Business School Press, Boston, MA.

Silk, Alvin J., Lisa R. Klein and Ernst R. Berndt. 2001. "The Emerging Position of the Internet as an Advertising Medium." *Netnomics* 3, 129:148.

Silk, Alvin J., Lisa R. Klein and Ernst R. Berndt. 2002. "Intermedia Substitutability and Market Demand by National Advertisers." *Review of Industrial Organization* 20: 323-348, 2002.

__________, Simon Cowan, and John Vickers. 1994. *Regulatory Reform—Economic Analysis and British Experience*. Cambridge, Massachusetts: MIT Press.

Small, John, and Julian Wright. 2000. "Decentralized Interchange Fees in Open Payment Networks: An Economic Analysis," mimeo, University of Aukland, December.

Stavins, Joanna. 1997. "A Comparison of Social Costs and Benefits of Paper Check Presentment and ECP Truncation," *New England Economic Review*, (July/August), pp. 27-44.

Stigler, George J., and James K. Kindahl (1970) *The Behavior of Industrial Prices*. New York: National Bureau of Economic Research.

Stiglitz, Joseph and Carl Walsh. 2002. *Principles of Microeconomics*. New York: W. W. Norton.

Swedish Bankers' Association. 2004. *Banks in Sweden—Facts about the Swedish banking market*. September.

Thaler, Richard H. and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Tirole, Jean. 1988. *The Theory of Industrial Organization*, Cambridge, Massachusetts: MIT Press.

Tyagi, Rajeev K. (1999) "A Characterization of Retailer Response to Manufacturer Trade Deals," *Journal of Marketing Research*, 36: 510-516.

United States Census Bureau. 2004. "2002 Economic Census: Industry Series," online: http://www.census.gov/econ/census02/guide/INDSUMM.HTM.

United States Department of Commerce: Bureau of Economic Analysis. 2003 "Imputations in the National Income and Product Accounts," September 29.

__________. 2004 "Gross Domestic Product: Implicit Price Deflator," November 30. van Raalte, Chris, and Harry Webers. 1998. "Spatial Competition with Intermediated Matching," *Journal of Economic Behavior and Organization*, vol. 34, no. 3 (March), pp. 477-488.

U.S. Department of Commerce. 2008. *Quarterly Retail E-Commerce Sales 3rd Quarter 2008*, Press Release (November 19).

Varian, Hal R. 2007. "Position Auctions." *International Journal of Industrial Organization* 25(6) Winter: 1163-1178.

"Verizon posts in-line 1Q adj eps; wireless biz adds 1.5M customers." 2008. *Forbes.com*. (April 28), at <http://www.forbes.com/feeds/afx/2008/04/28/afx4939948.html>.

Vickers, John. 2005. "Public Policy and the Invisible Price: Competition Law, Regulation, and the Interchange Fee, paper presented at a Conference *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005

Visa International. 2001. *International Country Overviews*, October.

Visa USA (various years) "Payment Systems Panel Study."

Viscusi, W. Kip, John M. Vernon, and Joseph E. Harrington, Jr. 2000. *Economics of Regulation and Antitrust*, Cambridge, Massachusetts: MIT Press.

Walters, Rockney G. (1989) "An Empirical Investigation into Retailer Response to Manufacturer Trade Promotions," *Journal of Retailing*, 65(2): 253-272.

Weiner, Stuart E., and Julian Wright. 2005. "Interchange Fees in Various Countries: Developments and Determinants," paper presented at a Conference *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005.

Weiner, Stuart E. and Julian Wright (2005) "Interchange Fees in Various Countries: Developments and Determinants," Working Paper, April 25.

Werden, G.J. and L. M. Froeb (1994). "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy," Journal of Law, Economics, & Organization, 194, pp. 407-426.

Willig, R. (1991). "Merger Analysis, Industrial Organization Theory, and Merger Guidelines," Brookings Papers on Economic Activity: Microeconomics, 281, pp. 292-293.

Winston, Clifford, and Steven A. Morrison. 1986. *The Economic Effects of Airline Deregulation*, Washington, DC: The Brookings Institution.

Wolters, Timothy (2000) "Carry Your Credit Card in Your Pocket: The Early History of the Credit Card at Bank of America and Chase Manhattan," *Enterprise and Society*, 1: 315-354.

Wolters, Timothy. 2000. ""Carry Your Credit in Your Pocket": The Early History of the Credit Card at Bank of America and Chase Manhattan," *Enterprise and Society*, vol. 1, no. 2 (June), pp. 315-354.

Wong, Wylie (1998) "Netscape Applauds Microsoft Suit," *TechWeb,* May 20, http://www.techweb.com/wire/story/msftdoj/TWB19980519S0007.

Wright, J. (2004). "The Determinants of Optimal Interchange Fees in Payment Systems," Journal of Industrial Economics, Vol. 52, No. 1, March, pp. 1-26.

Wright, Julian. 2000. "An Economic Analysis of a Card Payment Network." Mimeo, NECG and University of Auckland, December.

___________. 2003a. "Optimal Card Payment Systems," *European Economic Review*, vol. 47, no. 4 (August), pp. 587-612.

___________. 2003b. "Pricing in Debit and Credit Card Schemes," *Economics Letters*, vol. 80, no. 3 (September), pp. 305-309.

___________. 2003c. "Why Do Firms Accept Credit Cards?" mimeo, University of Auckland.

___________. 2004a. "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, vol. 52, no. 1 (March), pp. 1-26.

___________. 2004b. "One-sided Logic in Two-sided Markets," *Review of Network Economics*, vol. 3, no. 1 (March), pp. 44-64.

Yahoo Press Release. 2007. *"Yahoo!'s New "SmartAds" Meld Brand and Direct Response Advertising*." (July 2), at <http://yahoo.client.shareholder.com/PRESS/ReleaseDetail.cfm?ReleaseID=252034>.

PLATFORM ECONOMICS:

Essays on Multi-Sided Businesses

David S. Evans, is an economist, business advisor and Founder of Market Platform Dynamics. He is a recognized global authority on the design and implementation of platform business strategies, a body of work that he pioneered and that was captured in Invisible Engines: How Software Platforms Drive Innovation and Transform Industries, a 2006 American Publishers Association best business book award winner. David has consulted with many of the leading companies in the world and has served as an advisor to their CEOs and Boards of Directors. A frequent keynote speaker, David is often called upon to lead thought provoking industry discussions and to share his insights on the success and failure of platform businesses worldwide. David is also an advisor to and investor in many emerging ventures.

Published by Competition Policy International (CPI)

CPI provides comprehensive resources and continuing education for the global antitrust and competition policy community. Created and managed by leaders in the competition policy community, CPI delivers timely com-mentary and analysis on antitrust and global competition policy matters through a variety of media and applications.

Visit: competitionpolicyinternational.com

BUSINESS / Economics / Finance