



VOLUME 7 | NUMBER 1 | SPRING 2011

Competition Policy International

The Intersection of
Consumer Protection
and Competition in the
New World of Privacy

Julie Brill

Ofcom's Approach and
Priorities for Consumer
Protection & Empower-
ment

**Andrea Coscelli
& Claudio Pollack**

Applications Want to be
Free: Privacy Against
Information

**Michael R. Hammock
& Paul H. Rubin**

Journalism, Compe-
tition, and the Digital
Transition

Matthew Bye & Oliver Bethell

The Antitrust
Economics of Free

David S. Evans



Edited by David S. Evans



From the Editor

This Spring 2011 issue, our 13th, marks an important turning point. When we started in 2005, while we could have dispensed with a print edition and published entirely online, we didn't think either our authors or readers were ready. Authors liked the prestige of print and many readers liked thumbing through a print journal. The spread of iPads together with drastic improvements in the technology for producing online periodicals convinced us late last year to embrace online media fully and to shed the constraints of a print edition. As you get used to the benefits of consuming insights on competition policy in this new format we hope you'll agree that it was the right decision.

Among the benefits of our new format, CPI can now include video and audio material in addition to traditional print. We make use of this capability to kick off this issue. Two of the most influential competition officials in the world were gracious enough to talk with me shortly before releasing this issue. In my interviews with Joaquin Almunia, the Commissioner in charge of competition policy for the European Union, and Jon Leibowitz, Chairman of the U.S. Federal Trade Commission, we discussed their priorities, philosophies, and even touched on personal interests.

Fitting with this adventure into new media the substantive focus of much of this issue concerns digital media and communications. We begin with a symposium including contributions that highlight the interplay among these new technologies, consumer privacy, and consumer protection. FTC Commissioner Julie Brill kicks off the discussion followed by contributions by Andrea Coscelli and Claudio Pollack of OfCom, the United Kingdom's communications regulator; Professors Michael Hammock and Paul Rubin; and Google lawyers Matthew Bye and Oliver Bethell.

The next three articles focus on antitrust issues that are important to the web economy. Many products and services in this new economy are "free" and, exercising the occasional privilege that I take in publishing in this journal, I examine what—if any—import a price of zero has for antitrust. Latham and Watkins lawyer Hanno Kaiser provides his take on a recent debate in the popular media asking "Is the Web Dead"—to quote *Wired*—in the context of closed versus open platforms. Lastly, Manish Agarwal and David Round from the Centre for Regulation and Market Analysis at the University of South

Australia provide a short history and analysis of a business that is generating much antitrust scrutiny these days: search engines.

The next three articles provide a bit of an intermission from the web. Jean Tirole from the Toulouse School of Economics looks at what economists have learned about the role and regulation of interchange fees—the system where merchants that take a card for payment pay fees indirectly to the bank that issued the card to the consumer—in payment systems. Former Pfizer lawyer and current Fordham Professor Kent Bernard examines the challenges competition authorities face when analyzing mergers in innovation markets, providing case studies in the pharmaceutical industry. Then John Temple Lang provides a survey on how EU law treats—and should treat—practices involving aftermarkets.

In another break from the web, two years after the financial crisis, Bruno Lassere, Président du Conseil de la concurrence in France, looks at lessons learned from a competition authority viewpoint, asking among other questions, “Is antitrust enforcement increasingly irrelevant?”

Angela Zhang returns us to our main theme with an analysis of the Chinese court’s decision in the *Baidu* case, involving the leading search engine in China. The case is important because it reflects a serious attempt by a Chinese judge, in a private antitrust case, to deal with alleged anticompetitive practices involving a business under scrutiny in many other jurisdictions: search engines. Cleary Gottlieb lawyer Zhang appeals to the two-sided literature to argue the court got much of the analysis wrong.

The Classic for this issue is by Jeff Rohlfs, then at Bell Laboratories, in which he laid the foundation for the modern work on network effects. MIT Professor Richard Schmalensee argues that this article was really before its time as it provides an excellent economic model of Facebook—now the most trafficked web site on the planet. Schmalensee argues that this classic is well worth re-reading because its insights are even more important today than when the article was published almost four decades ago.

As always, on behalf of the competition policy community, we thank all the men and women who contributed to this issue.

David S. Evans
University of Chicago and University College London



*Interviews
with Antitrust
Leaders*

Interviews with Antitrust Leaders

With David S. Evans¹



Competition policy has grown explosively in the last quarter century. There are more than 100 countries with competition authorities and many of these authorities are becoming more active, influential, and professional. They are also talking with each other on a regular basis through many formal and informal channels. The managers of these authorities are assuming increasing responsibilities and wearing many hats—not only as regulators but also as advocates of markets. Beginning with this issue, Competition Policy International, in its publications and online, will be showcasing competition authorities through articles, interviews, and other formats.

To begin, we present interviews, conducted in Spring, 2011, with two of the world's most influential antitrust regulators, FTC Chairman Jon Leibowitz and European Commission Vice President Joaquín Almunia.

¹David S. Evans is Chairman, Global Economics Group; Lecturer, University of Chicago Law School; Executive Director, Jevons Institute for Competition Law and Economics; and Visiting Professor, University College London. He can be contacted at devans@globaleconomicsgroup.com.

I. Interview with Jon Leibowitz, Commissioner, Federal Trade Commission

Jon Leibowitz was sworn in as a Commissioner on September 3, 2004, and designated to serve as Chairman of the Federal Trade Commission on March 2, 2009, by President Barack H. Obama.



During his tenure as Chairman, Jon has focused on stopping scams that prey upon consumers suffering from the economic downturn; preserving competition in healthcare and blocking anticompetitive “pay-for-delay” patent settlements in the pharmaceutical industry; promoting competition and innovation in the technology sector through law enforcement and policy initiatives; and protecting consumers’ privacy—especially while they are using the Internet.

In this interview, Jon discusses a wide range of topics—from privacy, to last-dollar frauds, to the Intel case—and even why he has rap music on his iPad.

Listen to David Evans’ interview with Chairman Leibowitz:

LISTEN

II. Interview with Joaquín Almunia, EU Commissioner, Directorate-General for Competition



On 27 November 2009, José Manuel Barroso, President of the European Commission, nominated Joaquín Almunia, to serve as the new Competition Commissioner.

Vice President Almunia noted, “As Competition commissioner, I’m here to ensure that competition policy delivers for consumers and for businesses. To that end I will focus on fighting against cartels, preventing dominant companies from abusing their market power in any sector or any country in Europe, and maintaining a rigorous scrutiny of proposed mergers. I also intend to further consider how to achieve effective compensation for victims of illegal antitrust behavior.”

In this interview, Jon discusses a wide range of topics—from cartel enforcement and level of fines, to dealing with rapidly changing markets—and even revealing his fondness for American blues music.

Listen to David Evans’ interview with Commissioner Almunia:

[LISTEN](#)



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

The Intersection of Consumer Protection and Competition in the New World of Privacy

Julie Brill

Federal Trade Commission

The Intersection of Consumer Protection and Competition in the New World of Privacy

*Julie Brill**

Privacy issues are becoming increasingly important during this time of rapid technological advance. This article addresses the important question of how the FTC might balance the consumer protection concerns arising in the context of privacy with competition issues. It will first examine the basic principles of consumer protection and competition law, the two core missions of the FTC, and then take a look at some cases and other actions by the FTC outside the privacy realm that illustrate the different modes of interaction between the two areas of law. The agency's careful balance of its two core missions becomes clear through this exercise. Next, the article will describe the most recent evolution of privacy law at the agency, and the FTC's preliminary staff report on privacy. Included in the discussion will be a review of some of the latest privacy protection proposals from industry members. Finally, the article will discuss the interplay of some core consumer protection and competition principles in analyzing the privacy protection proposals.

*Commissioner, U.S. Federal Trade Commission. The views expressed in this article are my own and do not necessarily reflect the views of the Commission or the other Commissioners. I am grateful to my attorney advisor, Holly Vedova, for her invaluable assistance in preparing this article.

I. Introduction

The last five years have seen tremendous change in the on-line world. Technological advances have allowed on-line companies to develop new means to rapidly collect and share consumer information for use by behavioral advertisers and others. These changes can benefit consumers: Behavioral advertisers use the information to create ads more closely targeted to consumers' interests, increasing revenue to content providers, thereby funding much of the free content on the internet. But the changes also raise significant consumer privacy issues. Consumers may be unaware of the extent to which their on-line habits are bought and sold or, if they are aware, some consumers may curtail economic or other activity out of fear of the consequences.

For the most part, the privacy issues that arise in this context are based on consumer protection law. The FTC's recently released preliminary staff report on privacy is primarily focused on consumer protection issues.¹ However, some of the issues that arise in the privacy realm could also present competition concerns. Thus privacy joins a number of other issues at the Federal Trade Commission involving both consumer protection and competition claims. The intersection of these two areas of law is of growing significance to the business community, consumers, and practitioners, as well as to regulators. Sometimes the principles at the heart of these two areas of law point to conflicting results, while at other times they work in harmony. As in other areas of the law, the consumer protection concerns arising in the context of privacy will need to be balanced with competition issues.

To help shed light on how the FTC might undertake this task, this article will first examine the basic principles of consumer protection and competition law, the two core missions of the FTC, and then take a look at some cases and other actions by the FTC outside the privacy realm that illustrate the different modes of interaction between the two areas of law. The agency's careful balance of its two core missions becomes clear through this exercise. Next, the article will describe the most recent evolution of privacy law at the agency, and the FTC's preliminary staff report on privacy. Included in the discussion will be a review of some of the latest privacy protection proposals from industry members. Finally, the article will discuss the interplay of some core consumer protection and competition principles in analyzing these proposals.

THUS PRIVACY JOINS
A NUMBER OF OTHER ISSUES
AT THE FEDERAL TRADE
COMMISSION INVOLVING BOTH
CONSUMER PROTECTION AND
COMPETITION CLAIMS.

II. Consumer Protection and Competition Laws Both Address Distortions in the Marketplace

Consumer protection and competition law share at least one core concept: protecting consumers by removing distortions in the marketplace. Often the under-

lying conduct prohibited by these two areas of law impacts consumers in different ways. Conduct prohibited by consumer protection law usually involves individual businesses acting in a way that has a direct impact on consumers, for example, by deceiving or misleading them through false or deceptive advertising. A prime example in the privacy area is the FTC's recent complaint and settlement against EchoMetrix, where the company sold software to enable parents to monitor their children on-line, but failed to adequately disclose that the software also collected information about the kids' on-line activities and then sold that information to third-party marketers.²

Conduct prohibited by competition law also affects consumers, but the impact may not be as direct as on the consumer protection side because the prohibited practices in the first instance affect competition between businesses which then impacts consumers, for example, in the form of higher prices. An example involving privacy could be a situation where an on-line platform provider with a dominant share of the market introduces a privacy protection program that severely disadvantages a competitor.

Notwithstanding this difference in consumer impact, as former FTC Commissioner Tom Leary has noted, it takes only a few more moments of thinking about consumer protection and competition law to understand that these two areas of law share the common goal of addressing distortions in the marketplace that are designed to increase, or have the effect of increasing, the sales and profitability of a business or an industry in a manner detrimental to consumers.³ Consumer protection law addresses distortions that take place on the demand side of the transaction: Consumers' choices in the marketplace are negatively impacted, for example, by deceptive advertising that gives consumers the false

CONSUMER PROTECTION LAW
ADDRESSES DISTORTIONS THAT
TAKE PLACE ON THE DEMAND
SIDE OF THE TRANSACTION:
COMPETITION LAW ADDRESSES
DISTORTIONS THAT TAKE
PLACE ON THE SUPPLY SIDE.

impression that a product or service is worth more than it really is. Competition law addresses distortions that take place on the supply side: Anticompetitive practices, for example, exclude competitors or restrict supply among competitors, thereby elevating prices.⁴

However, as with most things in the real world, the distinction is not always so neat. Occasionally competition law addresses distortions that take place on the demand side; for example, when challenging anti-competitive practices that increase consumer switching costs. On these occasions competition law is even more closely aligned with consumer protection law because the competition law focuses on demand side conduct that decreases consumer choice or autonomy. It is easy to see how this could come into play in the privacy area. As on-line firms develop new privacy protections, one result could be increased consumer switching costs, something disfavored by competition law principles in certain situations. Similarly, consumer protection law occasionally addresses conduct aimed at competitors—for example, deceptive practices tar-

getting the perceived performance of competitor products—which, in turn, harms consumers. This conduct could arise in the privacy arena where, for example, a vertically integrated platform provider discriminates in the privacy protections it offers based on whether a competitor’s website is implicated. Such action could raise concerns about misrepresentation and deception.

The analysis becomes even more interesting when the conduct-distorting commerce implicates *both* consumer protection and competition principles. In those situations, the analysis is not as simple as in the above examples. As discussed below, tensions between the principles promoted by each area of law can arise; however, there are also instances when the two areas of law work in harmony.

III. Tension Between Competition and Consumer Protection Laws

Sometimes the principles promoted by competition law have the potential to trump consumer protection concerns. The *California Dental* case is an interesting example of the circumstances under which competition concerns can override facially legitimate consumer protection concerns.⁵ There, the FTC challenged a dental association’s ethical code that governed competing dentists’ advertisements of the price, quality, and availability of their services. The association’s ethical code prohibited its dentist members from making claims of across-the-board discounts off the dentists’ regular prices for certain groups of patients, such as senior citizens.⁶

The dental association claimed that the restrictions were needed because, even though some of the ads truthfully described the dentists’ fees, the association was concerned that the ads could not adequately disclose all the variables related to the fees, rendering the ads potentially misleading. Officials of the association testified that, in determining whether a particular ad was in violation of the code, they would attempt to determine whether the ad in its entirety would be misleading to a prudent person.⁷

Superficially, the prohibitions seemed consistent with consumer protection objectives. But the Commission concluded that, as enforced, the code was anti-competitive because it effectively prohibited even accurate advertising of prices and quality and restricted broad categories of advertising claims, without distinguishing between those that were deceptive and those that were not.⁸ As such, the code impaired dentists’ ability to engage in price competition. Thus, the Commission viewed its enforcement action as ensuring that practices aimed at promoting consumer protection objectives did not violate antitrust principles.

The Ninth Circuit essentially upheld the Commission’s opinion,⁹ but the Supreme Court concluded that the Ninth Circuit used a standard for analyzing the advertising restrictions—a “quick look” rule of reason analysis—that was too

abbreviated under the circumstances, and remanded the case for further proceedings.¹⁰ The Court did not say the restrictions had to be examined under a full-blown rule of reason (which would require the FTC to define a market and demonstrate that the association had market power). Rather, the Court simply said that the justifications for the restraints were sufficiently substantive that “[f]or now, at least, a less quick look was required.”¹¹ The Court based its ruling on a belief that the advertising restrictions could have had a net pro-competitive effect on competition, or no effect at all, and that the restrictions were, on their face, designed to avoid false and deceptive advertising, something particularly

AT A MINIMUM, BEFORE
COMPETITION PRINCIPLES CAN
TRUMP CONSUMER PROTECTION
CONCERNS, ANY LEGITIMATE
CONSUMER PROTECTION ISSUES
MUST BE IDENTIFIED AND
BALANCED AGAINST THE
COMPETITIVE HARM.

important in a market characterized by disparities in the information available as between dentists and their patients.¹²

An important lesson to be drawn from the *California Dental* case is that it is not always easy to strike the right balance between competition and consumer protection concerns. A reasonable interpretation of the Supreme Court’s ruling would be that the Commission’s analysis of the association’s activities did not strike the appropriate balance between competition and consumer protection interests. While misuse of consumer protection objectives can clearly lead to liability under the competition laws, the Commission was not adequately sensitive to the consumer protection aspects of the underlying conduct. This raises the obvious question of what the appropriate level of legal scrutiny should be in matters where consumer protection is asserted as a justification for conduct that encroaches on competition concerns. At a minimum, before competition principles can trump consumer protection concerns, any legitimate consumer protection issues must be identified and balanced against the competitive harm.

IV. Industry Self-Regulation

The fact that the *California Dental* case involved a self-regulatory body is an important aspect of the competitive analysis and judicial decision. The California Dental Association is a very large professional association composed of competing members engaging in self-regulation.¹³ Industry self-regulation can be a very good thing, as it may be the most efficient way for an industry to police itself by combating fraud and protecting consumers. In most circumstances, industry self-regulation should be encouraged. For instance, in the privacy context, behavioral advertisers’ ubiquitous collection of consumer data without consumers’ knowledge prompted the FTC in 2009 to urge the on-line industry to develop a self-regulatory response.¹⁴ Of course, industry self-regulation may be too slow to develop, or inadequate in its provisions or reach, to effectively address consumer harms. The industry’s response at the time of the FTC’s 2009

call for self regulation of the privacy issues surrounding behavioral advertising was slow and inadequate.¹⁵

In addition, industry self-regulation may heighten concerns about harm to competition among members of the profession or trade. When competitors form a trade association to self-regulate, and collectively have a dominant position in the marketplace, the risk of competitive concerns grows, and the conduct must be closely examined. In *California Dental*, the fact that the association's members accounted for 75 percent of practicing dentists in California bolstered the Commission's competition concerns.¹⁶

Industry self-regulation also may further entrench some competitors' positions. Notably, the Commission recently brought several cases involving professional licensing boards that issued rules under the auspices of consumer protection, but which the Commission alleged harmed competition and consumers by reducing competitive alternatives. For example, in 2007 the Commission settled a case against the South Carolina Board of Dentistry involving the board's newly imposed requirement that a dentist examine every child before a dental hygienist could provide preventive care, such as cleanings, in schools.¹⁷ The rule prohibited the previously common practice of using dental hygienists as an alternative to dentists in certain settings such as schools. The Commission found that the rule led to fewer children receiving preventive dental care. The rule was particularly egregious in the Commission's view because it largely affected economically disadvantaged children.¹⁸ In a more recent case, the Commission filed an adjudicative complaint against the North Carolina Dental Board for taking actions to block non-dentists from providing teeth whitening services.¹⁹

In both of these cases, the dental boards argued that their rules were needed to prevent physical harm to consumers from non-dentists; an objective ostensibly grounded in consumer protection concerns. But the Commission's pursuit of both cases struck a different balance between consumer protection and competition concerns. In both cases, the Commission believed that the boards were using a consumer protection rationale as a pretext for their desire to limit competition from non-dentists.

V. Consumer Protection Requirements May Outweigh Concerns about Entry Barriers

In a variety of other important matters, consumer protection principles often take precedence over competition principles. For example, consumer protection principles may have the effect of limiting entry into markets by new firms and products, even though entry traditionally plays an important role in addressing competition concerns. This phenomenon can be seen, for example, with respect to advertising substantiation in the food industry. New food products introduced

in a market usually are heavily advertised to gain consumer awareness. But new entrants can get into trouble if, for example, their advertising contains health claims that are not substantiated. The Commission imposes a fairly rigorous substantiation standard for health or safety claims in food products. These claims must be supported by competent and reliable scientific evidence.²⁰

Complying with substantiation requirements may place a greater burden on new entrants because performing the scientific studies necessary to substantiate some claims may require significant resources. Some potential entrants might therefore find substantiation requirements to be significant entry barriers. However, in the context of advertising health claims about food, entry through unsubstantiated claims should not be considered legitimate entry. Thus, the need to comply with substantiation requirements should trump the competition objective of reducing barriers to entry.

The Commission's Endorsements and Testimonials Guides might be said to pose entry barriers as well.²¹ The Guides set forth important principles of truth-in-advertising. For example, an advertisement featuring a consumer claiming or implying that her experience with a product is "typical," when that is not the case, should clearly and conspicuously disclose the typical consumer experience.²² Similarly, the Guides state that ads featuring statements by endorsers who have been paid to sing the praises of a product should disclose the payment.²³

The principles underlying the Endorsements and Testimonials Guides could constrain the very type of advertising required for new market entrants to gain market share. Indeed, the Guides apply to advertising through bloggers and other social media, among the lowest cost forms of advertising available to new market entrants. The Guides therefore arguably make it harder for new entrants to gain market share through creative on-line advertising. Yet once again consumer protection principles supporting full disclosure about testimonials and endorsements should trump these potential competition concerns about entry.

IN CALIFORNIA DENTAL,
FOR EXAMPLE, THE POTENTIAL
FOR HARM TO COMPETITION
WAS STRONG, BUT THE CONSUMER
PROTECTION CONCERNS
WERE ALSO STRONG.

The balance between consumer protection and competition concerns seems fairly easy in these examples. Unsubstantiated health claims and false testimonials have obvious harmful effects on consumers. But in other situations, it can be more difficult to make the right call. In *California Dental*, for example, the potential for harm to competition was strong, but the consumer protection concerns were also strong. The fact that the Supreme Court was more influenced by the consumer protection aspects of the conduct than both the Commission and the Ninth Circuit shows that in some matters where the two principles pull in opposite directions, finding the right balance can be challenging.

VI. Where Consumer Protection and Competition Concerns Harmonize Towards the Same Result

In other matters, consumer protection and competition principles converge and mutually support each other in the analysis of conduct harmful to consumers. One area that has received close attention for possible anticompetitive conduct involves high-tech markets where firms appear to be attaining dominance. In this area, consumer protection problems can be intermixed with exclusionary conduct. The FTC's recent *Intel* case is a good example.²⁴

INTEL

In August 2010, the Commission settled its administrative adjudication against Intel, a case that alleged both competition and consumer protection law violations. The Commission alleged that, since 1999, Intel had unlawfully maintained a monopoly in the market for central processing units ("CPUs"), and sought to acquire a second monopoly in graphics processing units ("GPUs"), using a variety of practices that violated antitrust laws as well as the competition and consumer protection prongs of Section 5 of the FTC Act.²⁵

The Complaint alleged that, in 1999 and again in 2003, Intel's competitors started to release products that were superior to Intel products in performance and quality, threatening Intel's monopoly. In response, Intel engaged in several practices that the Commission believed were designed to block or slow down the adoption of competitive products and allow Intel to maintain its monopoly, all to the detriment of consumers.

The practices that raised consumer protection aspects of the case involved Intel's compiler.²⁶ Beginning in 2003, Intel introduced a new version of its compiler shortly before its competitor, AMD, released its technologically superior CPU. Intel's new compiler slowed the performance of software on AMD's CPU. The Commission believed that Intel failed to adequately disclose that the changes it had made to its compilers beginning in 2003 were the cause of the slower performance of AMD's CPU.²⁷ The Commission also believed that Intel intentionally misrepresented the cause of and potential solutions to the performance differences, in an effort to portray its competitor's product as inferior.²⁸

THE COMBINED COMPETITION AND CONSUMER PROTECTION VIOLATIONS IN INTEL WERE ENFORCED IN A HARMONIOUS MANNER TO PROTECT CONSUMERS.

The Commission's Consent Order puts Intel under important restrictions that will improve the competitive landscape for the CPU and GPU markets. The Order also contains equally important requirements traditionally employed in a consumer protection context, including requiring Intel to engage in corrective

advertising about its compilers, and to reimburse software developers and vendors harmed by Intel's allegedly deceptive conduct.²⁹

The combined competition and consumer protection violations in *Intel* were enforced in a harmonious manner to protect consumers. The Commission's ability to protect competition and consumer protection simultaneously in this case was facilitated by the fact that Intel aimed its allegedly anticompetitive conduct and its allegedly deceptive conduct at the same target on the supply side of the equation: its competitors.

VII. The Commission's Preliminary Privacy Report and Industry's Response

As noted at the outset, recent developments in on-line and off-line data collection have prompted substantial activity at the FTC in the last few years centered on privacy concerns. Some of the meatiest privacy concerns raise issues that fall squarely at the intersection of consumer protection and competition law, implicating many of the different modes of interaction between the two areas of law discussed above.

Privacy is a central element of the Commission's consumer protection mission. In recent years, advances in technology have made it possible for detailed information about consumers to be stored, sold, shared, aggregated, and used more easily and cheaply than ever, in ways not feasible, or even conceivable, before. These advances in technology have, among other things, allowed on-line companies to engage in behavioral, or targeted, advertising. As noted above, targeted advertising has many important benefits. Consumers receive information about products and services in which they are more likely to be interested.

Businesses can better target their advertising dollars to reach the right audience. Perhaps most importantly, this type of advertising supports a great deal of the internet's free access to rich sources of information.

THE DRAMATIC CHANGES
IN TECHNOLOGY HAVE
CHALLENGED THE VITALITY
OF THE COMMISSION'S
TRADITIONAL PRIVACY MODELS.

Yet serious privacy concerns arise when companies can easily collect, combine, and use so much information from and about consumers. The dramatic changes in technology have challenged the vitality of the Commission's traditional privacy models. As the report notes, it is hardly a surprise to discover that there are significant gaps in older privacy protection models. In the mid-1990s, the fair information practices model was prevalent, with its call for businesses to provide consumers with notice and choice about how their personally identifiable information is used.³⁰ Then in the early 2000s, the Commission and others shifted to a harm-based model, under which the regulatory framework focused on data security, data breaches, and identity theft.³¹

There are significant problems with each of these frameworks. The “notice and choice” model, as it is being used today, places too great a burden on consumers. Many notices are written in “legalese” and are therefore difficult for consumers to read. And delivery of notices on mobile devices with their smaller screens compounds these problems.

THE “NOTICE AND CHOICE”
MODEL, AS IT IS BEING USED
TODAY, PLACES TOO GREAT A
BURDEN ON CONSUMERS.

On the other hand, a harm-based model may not sufficiently address the myriad of harms that can result from insufficient privacy protections surrounding information about medical conditions, children, and sexual orientation, to name a few salient examples. The “harm” model is also fundamentally reactive: it addresses and corrects privacy and data security breaches after they have been discovered, rather than focusing on creating a climate in which privacy is part of the fundamental design of products and services being offered.³²

And both models focus on “personally identifiable” information, a concept which may be out of touch with technological advances that allow previously non-identifiable data to be “re-identified” with a consumer.³³

After grappling with these issues over the past year and a half, FTC staff issued a preliminary report for policymakers like Congress, as well as for industry, that proposes a framework for rethinking their approach to privacy.³⁴ The proposed framework urges both policymakers and the industry toward a more dynamic approach to addressing privacy in today’s technologically advanced landscape.

The main elements of the framework in the preliminary staff report include the following:

1. Companies should adopt a “privacy by design”³⁵ approach that involves building privacy protections into their everyday business practices, such as providing reasonable security for consumer data, collecting only the data needed for a specific business purpose, retaining data only as long as necessary to fulfill that purpose, safely disposing of data no longer in use, and implementing reasonable procedures to promote data accuracy.
2. Companies should improve the transparency of their data practices, including improving their privacy notices so that consumer groups, regulators, and others can compare data practices and choices across companies.
3. Companies should provide information to consumers about their data practices through simpler, more streamlined choices than have been used in the past. Choices should be clearly and concisely described, and offered at a time and in a context in which the consumer is making a decision about his or her data. The FTC took no position on

opt-in or opt-out in the report, but rather focused on whether the notice and choice mechanism offered is robust.

There are several different mechanisms that can be employed to provide more meaningful choices to consumers. The 2010 Report discusses a “Do Not Track” choice as one means of allowing consumers to exercise choice about collection and use of information about their on-line activity.³⁶

To implement Do Not Track, the 2010 Report indicates that the most practical method of providing consumer choice may be a browser-based approach.³⁷ This approach allows consumers to make persistent choices that travel with them through cyberspace, communicating their tracking preferences to every website they visit, giving consumers meaningful control over the information they share and the sort of targeted ads they receive. The staff report indicates

that other proposals besides a browser-based approach can work as well³⁸ and seeks input from commenters about other proposals.³⁹

THE 2010 REPORT'S
RECOMMENDATION OF A
DO NOT TRACK MECHANISM HAS
IGNITED A HEARTY RESPONSE.

The 2010 Report's recommendation of a Do Not Track mechanism has ignited a hearty response. Several proposals have been put forward by major industry players. Some are browser-based,⁴⁰ and others employ use of icons and cookies for consumers to express their tracking preferences.⁴¹ These proposals, and others, are rapidly developing.

To determine how successful any particular mechanism is in reaching the consumer protection goal of providing simplified choice, the Commission and others should examine the mechanism based on the following criteria, among others:

- (i) Is the mechanism easy to use? Or will the mechanism lead to multiple systems that can lead to consumer confusion?
- (ii) Is it universal? That is, is there participation by the vast majority of advertisers, ad networks, service providers and other relevant industry players?
- (iii) Does it provide for opt-out of collection of information, in addition to opt-out of use of the information for particular purposes, such as targeted advertising?
- (iv) Does the mechanism allow for consumer choice that is persistent?
- (v) Is the mechanism effective and enforceable?⁴²

Each of the Do Not Track mechanisms proposed by industry to date satisfies a different bundle of these criteria, and is therefore capable of fulfilling the consumer protection goals of a simplified choice mechanism to a greater or lesser extent than other mechanisms. As many of these proposals are still in

their early stages, it is as of this writing still too early to definitively opine about which of them will ultimately be the most successful in fulfilling the consumer protection mandate.

VIII. Competition and Consumer Protection with Respect to Privacy

Recently, several commentators have urged firms to compete based on how they collect, use, store, and dispose of consumers' information—that is, to engage in competition based on privacy.⁴³ This form of competition is clearly in its gestational stage. Some of the positive developments with respect to competition as it concerns privacy include companies' efforts to improve baseline data security standards for cloud computing services and to improve use of encryption by default for email service;⁴⁴ major on-line search engines' efforts to shorten retention periods for search data;⁴⁵ and development of new tools, including the Do Not Track mechanisms described above, that allow consumers to control their receipt of targeted advertisements and to see and correct the information companies collect about them for targeted advertising.⁴⁶

Any new framework for privacy should promote both competition and consumer protection principles. Encouraging “privacy by design” and other new ways of thinking about privacy may present firms with greater incentives to compete on privacy, thereby increasing consumer choice and opportunities in this area. In this way, both areas of law could be aligned to address demand side distortions of the marketplace.

Yet it is worthwhile to consider the precise contours of the alignment between competition and consumer protection concerns with respect to privacy. The 2010 Report's proposed new privacy framework arguably could raise concerns about the ability of new firms to enter a market. Some observers may ask: Can new firms design the kinds of dynamic, just-in-time notices that should now be provided? Can they adequately address concerns about personally identifiable information, secondary uses of information, and use of so-called “legacy” data collected under prior privacy regimes? Or will these new recommendations create a barrier to entry in markets that have been the hallmark of dynamism in our economy?

Rather than viewing the proposed new privacy framework as imposing potential barriers to market entry for new firms, the new framework might instead present market entrants with an *advantage*, by providing them with a guidepost for

ENCOURAGING “PRIVACY BY DESIGN” AND OTHER NEW WAYS OF THINKING ABOUT PRIVACY MAY PRESENT FIRMS WITH GREATER INCENTIVES TO COMPETE ON PRIVACY, THEREBY INCREASING CONSUMER CHOICE AND OPPORTUNITIES IN THIS AREA.

creating business models that address privacy concerns from the outset, rather than as an afterthought. Indeed, some more established data brokers and other information firms believe it is much easier for their newer competitors to design privacy protections into their new business models and new forms of consumer communications than it is to retrofit old systems to meet the realities of today's privacy concerns. New firms may well have a "leg up" on existing players, if they implement these recommendations at the start of their business endeavors.

In addition, with respect to both new and existing firms, the proposed new framework's principles may move regulators and businesses away from a reactive model that focuses on privacy concerns after harm is done and towards a model where companies are encouraged to entice consumers to use their products and services based, in part, on their privacy practices.

THE VARIOUS DO NOT TRACK PROPOSALS, PARTICULARLY SINCE THEY ARISE IN A SELF-REGULATORY CONTEXT, ALSO RAISE SOME INTERESTING ISSUES WITH RESPECT TO THE ALIGNMENT OF CONSUMER PROTECTION AND COMPETITION PRINCIPLES.

The various Do Not Track proposals, particularly since they arise in a self-regulatory context, also raise some interesting issues with respect to the alignment of consumer protection and competition principles. First, there could be competition concerns if a particular proposal disadvantages competitors of the platform offering the proposal, especially if the platform operator has a dominant market share and is vertically integrated. Depending on the circumstances, the proposal could result in an exclusionary practice similar to those addressed by the Commission in the *Intel* matter if competitors are blocked, or entry barriers are otherwise significantly raised.

Second, to the extent a proposal is offered by a group of competitors (for example, a trade association), there could be concerns if the competitors act in ways that favor their own economic interests to the detriment of other competitors and consumers. This is especially a problem if the trade association developing the proposal has a dominant share of the market, as in the *California Dental* and professional licensing board cases discussed above. As we have seen, oftentimes competitors can favor their own economic interests in the context of industry self-regulation undertaken in the name of "consumer protection," to the detriment of competition and consumers because the self-regulation reduces competitive alternatives.

In cases where industry implements a Do Not Track mechanism and other aspects of the new proposed privacy framework under the auspices of self-regulation, the Commission will need to watch developments closely, to ensure that such requirements, ostensibly aimed at protecting privacy, are not simply a means to keep out new entrants. As noted earlier, in other Commission actions involving self-regulatory regimes, there may be a tipping point at which self-regulation turns anticompetitive, particularly in cases where the mechanisms are developed by a trade association or industry players that have a dominant market position.

At the same time, the Commission and other policy makers must keep an eye on the consumer protection objectives for a Do Not Track mechanism. Unless a proposal is put forth by a firm with a significant presence in the market, or is adopted by a group of firms that together have a significant presence, there may be too many different proposals, creating consumer confusion. And with respect to proposals offered by an industry group that require adherence by industry members, unless all or the vast majority of industry members agree to abide by the proposal, it may fail to meet the universality criteria. Tensions like these, between consumer protection goals and competition issues, may arise, and will have to be carefully balanced as various industry and regulatory proposals are fleshed out.

IX. Conclusion

The latest developments in the fast-changing world of data collection and use raise many questions at the intersection of consumer protection and competition law. Some have easy answers, others do not. This article has covered some issues worth considering in this area and suggested some ways to analyze the issues. There will undoubtedly be further developments in the very near future as this dynamic industry continues to evolve, so policymakers and practitioners should keep a close watch on this space. It is undoubtedly the case that the Commission, with its unique focus on both consumer protection and competition law, will continue to take a strong interest in developments in privacy protection and the challenges these developments will present at the intersection of these areas of the law in the future. ▼

-
- 1 Fed. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICYMAKERS (2010) (preliminary FTC staff report) [hereinafter 2010 PRIVACY REPORT], available at <http://www.ftc.gov/os/2010/12/101201privacyreport.pdf>.
 - 2 *FTC v. Echometrix, Inc.*, No. CV10-5516 (E.D.N.Y. Nov. 30, 2010) (consent order).
 - 3 See Thomas B. Leary, *Competition Law and Consumer Protection Law: Two Wings of the Same House*, 72 ANTITRUST L.J. 1147, 1147-48 (2005).
 - 4 *Id.*
 - 5 *California Dental Ass'n v. FTC*, 526 U.S. 756 (1999).
 - 6 *California Dental Ass'n*, 121 F.T.C. 190, 192-94 (1996), *aff'd sub nom. California Dental Ass'n v. FTC*, 128 F.3d 720 (9th Cir. 1997), *vacated*, 526 U.S. 756 (1999).
 - 7 *Id.* at 347-48 (Azcueña, Comm'r, dissenting).
 - 8 *Id.* at 300-03, 307.
 - 9 See *California Dental Ass'n v. FTC*, 128 F.3d 720, 726-30 (9th Cir. 1997), *vacated*, 526 U.S. 756 (1999).
 - 10 See *California Dental*, 526 U.S. at 769-81.

- 11 *Id.* at 781.
- 12 *Id.* at 771-72; see also *id.* at 774-76, 778.
- 13 *Id.* at 759-62.
- 14 *FTC Staff Report: Self-Regulatory Principles for Online Behavioral Advertising* (2009), available at <http://www.ftc.gov/os/2009/02/P0085400behavadreport.pdf>.
- 15 Julie Brill, *Keynote Address to International Association of Privacy Professionals, The FTC and Consumer Privacy Protection* (December 7, 2010), available at <http://www.ftc.gov/speeches/brill/101207iapp.pdf>.
- 16 See *California Dental Ass'n*, 121 F.T.C. 190, 314-16 (1996), *aff'd sub nom.* *California Dental Ass'n v. FTC*, 128 F.3d 720 (9th Cir. 1997), *vacated*, 526 U.S. 756 (1999).
- 17 Press Release, Fed. Trade Comm'n, *South Carolina Board of Dentistry Settles Charges That it Restrained Competition in the Provision of Preventive Care by Dental Hygienists* (June 20, 2007), available at <http://www.ftc.gov/opa/2007/06/dentists.shtm>.
- 18 Opinion and Order of the Commission at 1-3, *South Carolina State Bd. of Dentistry*, Dkt. No. 9311 (F.T.C. July 30, 2004), available at <http://www.ftc.gov/os/adjpro/d9311/040728commissionopinion.pdf>.
- 19 Press Release, Fed. Trade Comm'n, *Federal Trade Commission Complaint Charges Conspiracy to Thwart Competition in Teeth-Whitening Services* (June 17, 2010), available at <http://www.ftc.gov/opa/2010/06/ncdental.shtm>; see also *Complaint, North Carolina Bd. of Dental Examiners*, Dkt. No. 9343 (F.T.C. filed June 17, 2010), available at <http://www.ftc.gov/os/adjpro/d9343/100617dentalexamcmt.pdf>. Note that the North Carolina Board of Dentistry matter is currently in litigation; the author did not participate in voting out the complaint due to her previous position at the North Carolina Attorney General's office.
- 20 See, e.g., *Order to Show Cause & Order Modifying Order, Kellogg Co.*, Dkt. No. C-4262 (F.T.C. May 28, 2010) (requiring "competent and reliable scientific evidence that is sufficient in quality and quantity based on standards generally accepted in the relevant scientific fields, when considered in light of the entire body of relevant and reliable scientific evidence, to substantiate that the representation is true," and defining "competent and reliable scientific evidence" to mean "tests, analyses, research, or studies that have been conducted and evaluated in an objective manner by qualified persons and are generally accepted in the profession to yield accurate and reliable results"), available at <http://www.ftc.gov/os/caselist/0823145/100602kelloggorder.pdf>; *Decision & Order, Indoor Tanning Ass'n*, Dkt. No. C-4290 (F.T.C. May 13, 2010) (same), available at <http://www.ftc.gov/os/caselist/0823159/100519tanningdo.pdf>.
- 21 *Guides Concerning the Use of Endorsements and Testimonials in Advertising*, 16 C.F.R. pt. 255 (2010).
- 22 16 C.F.R. § 255.2(b) (2010).
- 23 16 C.F.R. § 255.5 (2010).
- 24 Press Release, Fed. Trade Comm'n, *FTC Settles Charges of Anticompetitive Conduct Against Intel* (Aug. 4, 2010), available at <http://www.ftc.gov/opa/2010/08/intel.shtm>; *Decision & Order, Intel Corp.*, Dkt. No. 9341 (F.T.C. Oct. 29, 2010), available at <http://www.ftc.gov/os/adjpro/d9341/101102inteldo.pdf>.
- 25 Section 5 of the FTC Act prohibits "[u]nfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce." 15 U.S.C. § 45(a)(1). The first part of

the statutory language (unfair methods of competition) covers competition violations, and the second part (unfair or deceptive acts or practices) covers consumer protection violations.

- 26 Decision & Order, Intel Corp., Dkt. No. 9341 (F.T.C. Oct. 29, 2010), *available at* <http://www.ftc.gov/os/adjpro/d9341/101102inteldo.pdf>, ¶19. A compiler is a tool used by developers to write software. The compiler translates the “source code” of programs into “object code” that can be run as software on consumers’ computers. *Id.* ¶157.
- 27 *Id.* ¶¶158-59.
- 28 *Id.* ¶160.
- 29 *Id.* at 14-17.
- 30 For example, Title V, Subtitle A, of the Gramm-Leach-Bliley (GLB) Act, 15 U.S.C. §§ 6801-6809, requires, among other things, that financial institutions provide their customers with annual notices explaining their privacy practices and their customers’ rights to limit some sharing of the customers’ information. For more detail, see Federal Trade Comm’n, Bureau of Consumer Protection Bus. Ctr., *In Brief: The Financial Privacy Requirements of the Gramm-Leach-Bliley Act*, *available at* <http://business.ftc.gov/documents/bus53-brief-financial-privacy-requirements-gramm-leach-bliley-act>.
- 31 The FTC’s Safeguards Rule, for example, sets forth standards that certain financial institutions must use in developing, implementing, and maintaining reasonable safeguards to protect the security, confidentiality, and integrity of any nonpublic personal information pertaining to their customers. See 16 C.F.R. pt. 314 (2010) (implementing GLB Act §§ 501(b), 505(b)(2), 15 U.S.C. §§ 6801(b), 6805(b)(2)). In addition, numerous states have enacted breach notification laws. See, e.g., Samuelson Law, Tech. & Public Pol’y Clinic, Univ. of Cal.–Berkeley Sch. of Law, *Security Breach Notification Laws: Views from Chief Security Officers*, App. A, at 40-46 (Dec. 2007), *available at* http://www.law.berkeley.edu/files/cso_study.pdf.
- Both the FTC and the states have brought numerous law enforcement actions in the data security context. See, e.g., Press Release, Fed. Trade Comm’n, Twitter Settles Charges that It Failed to Protect Consumers’ Personal Information; Company Will Establish Independently Audited Information Security Program (June 24, 2010), *available at* <http://www.ftc.gov/opa/2010/06/twitter.shtm>; Agreement Containing Consent Order, Twitter, Inc., File No. 092-3093 (F.T.C. June 24, 2010), *available at* <http://www.ftc.gov/os/caselist/0923093/100624twitteragree.pdf>; Stipulated Final Judgment and Order for Permanent Injunction and Other Equitable Relief as to Defendants LifeLock and Davis, FTC v. LifeLock, Inc., No. 2:10-CV-00530-NVW (D. Ariz. Mar. 15, 2010) (ECF No. 9); Press Release, Office of the Atty. Gen., Cal. Dep’t of Justice, Brown Stops LifeLock from Misleading Consumers about Identity Theft Protection Services (Mar. 9, 2010) (announcing settlement entered into jointly with FTC and 34 other states), *available at* <http://ag.ca.gov/newsalerts/release.php?id=1869>; Assurance of Discontinuance, State v. TJX Cos., No. 465-6-09 WNCV (Vt. Super. Ct. June 23, 2009) (settlement entered into with 40 states and District of Columbia), *available at* <http://www.atg.state.vt.us/assets/files/TJX%20AOD.pdf>; Decision & Order, TJX Cos., Dkt. No. C-4227 (F.T.C. July 29, 2008) (consent order), *available at* <http://www.ftc.gov/os/caselist/0723055/080801tjxdo.pdf>; Press Release, Office of the Atty. Gen., State of Vt., Attorney General Sorrell Reaches Agreement with ChoicePoint (May 31, 2007), *available at* <http://www.atg.state.vt.us/news/attorney-general-sorrell-reaches-agreement-with-choicepoint.php>; Stipulated Final Judgment and Order for Civil Penalties, Permanent Injunction and Other Equitable Relief, United States v. ChoicePoint, Inc., No. 1:06-CV-0198-JTC (N.D. Ga. Feb. 15, 2006) (ECF No. 5); Decision & Order, BJ’s Wholesale Club, Inc., Dkt. No. C-4148 (F.T.C. Sept. 20, 2005) (consent order), *available at* <http://www.ftc.gov/os/caselist/0423160/092305do0423160.pdf>; Stipulated Consent Agreement & Final Order, FTC v. Toysmart.com, LLC, No. 00-11341-RGS, 2000 WL 34016434 (D. Mass. July 21, 2000).
- 32 See Daniel J. Solove, *Identity Theft, Privacy, and the Architecture of Vulnerability*, 54 HASTINGS L.J. 1227, 1232-46 (2003).

- 33 The Commission has been concerned about reidentification of previously non-identifiable data. See Letter from Maneesha Mithal, Assoc. Dir., Div. of Privacy and Identity Prot., FTC, to Reed Freeman, Morrison & Foerster LLP, Counsel for Netflix (Mar. 12, 2010), available at <http://www.ftc.gov/os/closings/100312netflixletter.pdf> (closing letter).
- 34 See 2010 PRIVACY REPORT.
- 35 Privacy By Design is an approach that Ann Cavoukian, Ph.D., Information and Privacy Commissioner of Ontario, has advocated. See Ann Cavoukian, Ph.D., Ont. Info. & Privacy Comm’r, *PbD: Privacy by Design*, available at <http://www.ipc.on.ca/images/Resources/privacybydesign.pdf>.
- 36 See 2010 Privacy Report at 63-69 (2010); Center for Democracy and Technology, *What Does “Do Not Track” Mean? A Scoping Proposal By the Center for Democracy and Technology* (January 31, 2011), available at <http://cdt.org/files/pdfs/CDT-DNT-Report.pdf>.
- 37 See 2010 Privacy Report at 66.
- 38 See *id.* at 63-64.
- 39 See *id.* at A-4.
- 40 See, e.g., Microsoft’s Internet Explorer 9 “Tracking Protection” technology (Press Release, Microsoft, Providing Windows Customers with More Choice and Control of Their Privacy Online with Internet Explorer 9 (Dec. 7, 2010), available at <http://www.microsoft.com/presspass/features/2010/dec10/12-07ie9privacyqa.mspx>); Mozilla and Stanford University Researchers’ Firefox browser Version 4 with its Do Not Track header feature (Mozilla Blog, Mozilla Firefox 4 Beta, now including “Do Not Track” capabilities (Feb. 8, 2011), available at <http://blog.mozilla.com/blog/2011/02/08/mozilla-firefox-4-beta-now-including-do-not-track-capabilities/>); and Apple’s Do Not Track feature to be included in the next version of its Safari browser (Nick Wingfield, *Apple Adds Do-Not-Track Tool to New Browser*, WALL S. J., April, 14, 2011, available at <http://online.wsj.com/article/SB10001424052748703551304576261272308358858.html>).
- 41 See the Digital AdvertisingAlliance’s proposal described at <http://www.aboutads.info>. The Digital Advertising Alliance is a coalition of marketing and business groups, including over 300 major ad networks (over 300), as well as the American Association of Advertising Agencies, the Direct Marketing Association, the Better Business Bureau, the Interactive Advertising Bureau, the American Advertising Federation, the Association of National Advertisers, and the Network Advertising Initiative. See also Google’s free, downloadable add-on for its Chrome browser, which provides a single, persistent opt-out of behavioral advertising by members of the Interactive Advertising Bureau, a self-regulatory group. Google Public Policy Blog, *Keep your opt-outs* (Jan. 24, 2011), available at <http://googlepublicpolicy.blogspot.com/2011/01/keep-your-opt-outs.html>.
- 42 Prepared Statement of the Federal Trade Commission on the State of Online Consumer Privacy before the Senate Committee on Commerce, Science and Transportation (March 16, 2011) at page 16, available at <http://www.ftc.gov/os/testimony/110316consumerprivacysenate.pdf>.
- 43 See Emily Gray, *FTC to Boost Competition in Privacy Protection*, GLOBAL COMPETITION REV., Sept. 23, 2010, available at <http://www.globalcompetitionreview.com/news/article/29065/ftc-boost-competition-privacy-protection-/>; Pamela Jones Harbour & Tara Isa Koslov, *Section 2 in a Web 2.0 World: An Expanded Vision of Relevant Product Markets*, 76 ANTITRUST L. J. 769, 773-75, 792-97 (2010).
- 44 See 2010 Privacy Report at 45, 47-48 & n.122 (2010).
- 45 See *id.* at 47-48 & n.122.
- 46 See *id.* at 63 & n.149.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Ofcom's Approach and Priorities for Consumer Protection & Empowerment

Andrea Coscelli & Claudio Pollack
OfCom

Ofcom's Approach and Priorities for Consumer Protection & Empowerment

*By Andrea Coscelli & Claudio Pollack**

This paper discusses Ofcom's current activities related to consumer protection and empowerment. It describes our approach and framework for analysis, and goes on to examine those areas we currently treat as our top priorities. We do so by exploring the following questions:

- What is the role of consumer policy?
- What do we mean by consumer protection and empowerment?
- What issues have given rise to concerns in our recent experience?
- What tools do we have to improve consumer outcomes, taking account of the impact of market mechanisms and the role of incentives?
- What is the evidence of the effectiveness of our approach to date?

We also briefly discuss our proposed intervention against the sale of automatically renewable contracts to purchasers of fixed voice and fixed broadband services.

Our protection and empowerment work complements our competition work as it addresses areas where markets without dominant providers are not functioning perfectly for consumers in terms of their ability to compare and switch providers easily as well as to negotiate, understand, and enforce contracts. As the OFT

*Andrea Coscelli is Director of Competition Economics and Claudio Pollack is Group Director, Consumer at Ofcom. The views expressed in this article are personal and they do not represent Ofcom's views.

states “markets work well when there are efficient interactions on both the demand (consumer) and supply (firm) side. On the demand side, confident consumers activate competition by making well-informed and well-reasoned decisions which reward those firms which best satisfy their needs.”¹

In addition to our powers using regulations and our work with industry, we also discuss the scope for incorporating greater use of comparative information and behavioral economics to improve market outcomes and reduce consumer harm.

I. What Is the Role of Consumer Policy?

Competition policy seeks to use regulatory instruments mainly to address supply-side market failures, especially as regards number of suppliers, market shares, and barriers to entry and exit, which, if left unchecked, could lead to one or more suppliers being in a position to exploit a dominant position. This, in turn, can result in poor consumer outcomes such as excessive or inefficient charges, and insufficient incentives to innovate and invest in new product offerings.

By contrast, consumer policy has often been referred to as the “demand side” of competition policy.² Standard economic theory suggests a number of conditions need to be satisfied if markets are to deliver efficient outcomes for consumers. Where these conditions are not present, this can give rise to market failures—where the market fails to secure efficient outcomes for consumers. While supply-side failures can result from barriers to entry and exit, for example, demand-side failures can result from imperfect information or high search (or switching) costs. Equally, sub-optimal outcomes are possible wherever consumers’ actual ability to engage with markets falls short of complete rationality. So a lack of information can result in consumers not getting the best from markets,

but so can consumers’ inability to absorb and process the information that does exist.

BY CONTRAST, CONSUMER POLICY HAS OFTEN BEEN REFERRED TO AS THE “DEMAND SIDE” OF COMPETITION POLICY.

The identification of a demand-side market failure does not, of itself, imply a need for regulatory action. Just as there are instances of market failure, there is also the possibility of regulatory

failure. A regulatory action intended to correct a market failure can create a burden, ultimately passed on to consumers, which exceeds the cost the intervention is seeking to address. It can also occur where the necessary imperfections of the analysis mean that the regulatory interventions result in unintended consequences that lead to harm that is greater than that which we, as the regulator, are seeking to correct. We need to set the bar for regulation, particularly for costly or risky interventions, at a relatively high level. In many cases, that is underlined by the legal framework and by the scrutiny our decisions receive from the relevant courts.

In addition, the market itself can be quite adept at finding market-based solutions to apparent market failures. For example, providers of low quality services have little incentive to reveal the quality of their services and this can give rise to insufficient or asymmetric information in the market. But the market has developed a number of remedies and proxies for this problem. For example, brands can provide consumers with a recognisable and easy-to-process proxy for quality. And, intermediaries have made it their business to step in to help fill informational gaps.

On a similar theme, we need to consider whether any demand-side market failure is an enduring feature or is transitory as this, too, will influence the desirabil-

ity of regulatory remedies. For example, a bad outcome resulting from complexity of products could merely be the feature that encourages learning by consumers and allows them to “catch up.”³ But if the nature of products continues to change, then it becomes less likely that consumers will “catch up” in this way. Again, behavioral economics provides valuable insights in this area.⁴ Finally, it is important to bear in mind that regulatory intervention is, at least to some extent, a substitute for consumer learning so the expected outcome absent intervention would be expected to change over time.

So in some cases Ofcom has decided not to act even where it accepts that there is evidence of sub-optimal outcomes for consumers. We will choose to do so when all the actions available either will not address the harm or will do so subject to adverse consequences which are in excess of the likely benefits. And this will always be a difficult choice for a regulator to make when facing demands for intervention by consumer groups or industry participants.

SO IN SOME CASES OFCOM HAS DECIDED THAT THE RISK OF REGULATORY FAILURE IS TOO HIGH EVEN WHERE IT ACCEPTS THAT THERE IS EVIDENCE OF SUB-OPTIMAL OUTCOMES FOR CONSUMERS.

The key features of the communications sector that can result in demand-side market failure include:

- **The relatively recent history of liberalisation**, which means some consumers are not familiar with the need, or do not have the skills, to shop around for the products and prices that most suit their preferences. This is reflected, for example, by low awareness of competitors in some segments of the population, particularly in fixed-voice telephony.
- The **complexity of services** and the rate of change in offerings, which means that consumers may find it difficult to understand and compare the attributes of the services they are considering buying in order to reach a suitable decision.⁵ This could, in turn, mean that individual consumers are not making the decisions they would make if they had access to, and could process, information perfectly. It also means that some consumers could find themselves locked into contracts that differ markedly from the product they thought they were purchasing. In the extreme, this feature can give rise to a particular risk of scams.
- **Pricing complexity** can also create difficulty for consumers. For example, in purchasing a mobile contract a consumer needs to be aware of their future consumption of the various attributes of the service, including different call types and the various “additional charges” they are faced with. This can make choosing the best deal very difficult.⁶
- **Switching (and number porting) processes**, left to industry, have led to very poor consumer experiences in some instances.⁷ These can

result in direct harm to those consumers that have attempted to switch, and can lead to a reduction in competitive intensity in the market if the result is a lowering in the propensity to switch. This has been a particular feature in fixed-voice and broadband markets. This is because these services often share the same infrastructure (Openreach's local loop). Coordination is needed for purely technical reasons and there are low and asymmetric incentives for providers to agree to processes that are good for consumers, either due to coordination failure or because some providers have low incentives to work towards lower switching barriers. That is, each change in switching processes will generate likely commercial "winners" and "losers."

- We are also seeing an increasing trend towards providers seeking to introduce **contractual restrictions** on customers' ability to switch. Some of these are clearly matched by corresponding consumer benefits. For example, where a provider absorbs an upfront cost (such as a handset subsidy) it appears appropriate to then bind the consumer into a proportionate minimum contract period. However, we have concerns that some restrictions to switching may not be justified by the corresponding customer benefit. Behavioral economics plays an important role in the analysis. If consumers were perfectly rational and there was perfect transparency, consumers could make an informed choice based on their assessment of the downside of future restrictions to switching relative to the benefits being offered in return. However, if consumers do not properly understand or evaluate the cost of future switching restrictions, then concerns may arise. We discuss in detail one such example in section VI below.
- In situations where both parties to a contract have equal information and resources, they will also have equal ability to negotiate terms and to enforce. However, in consumer-to-business contracts this is not the case. **Terms are standard** and it is often **difficult for consumers to understand** the implications of **small print**. Standard court routes for enforcement may be prohibitively costly where disputed values are small. Absent regulatory action, the results can be that providers slip in terms that perfectly rational and informed consumers would not have consented to, or that breaches to terms go unchallenged.
- The growth of **bundles** can generate significant benefits for consumers, but can also serve to exacerbate many of the issues identified above.⁸

II. What Do We Mean by Consumer Protection and Empowerment?

Our starting point is that competition is the best means of delivering good outcomes for consumers. This is at the heart of Ofcom's regulatory principles and is derived from the statutes.⁹ Regulatory action designed to improve consumer out-

comes does so by seeking to enhance the ability of the market to deliver good outcomes for consumers. Only where that is not possible, we will look for mechanisms that deliver those outcomes that the market will not deliver. For example, we have resisted pressure to introduce detailed regulation of competitive providers' customer service standards, looking instead at informational remedies that would encourage providers to invest in the quality of their customer services where there is a consumer demand for such improvements.

We use the term “consumer protection” to describe those actions we take that lead to consumers being more directly protected from scams and unfair practices. The term relates to those things we consider providers can and cannot “do” to consumers. For example, in response to a growth in public concern and evidence of harm from silent calls we introduced new rules to limit the behavior of call centers. We fell short of banning the dialling equipment that gives rise to abandoned calls, but we did specify rules that were intended to dramatically reduce the harm resulting from their use.¹⁰

WE USE THE TERM “CONSUMER PROTECTION” TO DESCRIBE THOSE ACTIONS WE TAKE THAT LEAD TO CONSUMERS BEING MORE DIRECTLY PROTECTED FROM SCAMS AND UNFAIR PRACTICES.

The critical categories of consumer protection concerns are:

- Practices that lure consumers into contracts or services they have not consented to;
- Harm from process problems, and those that occur once service is being provided; and
- Practices which make it hard for consumers to exit the contract.

We use the term “consumer empowerment” to refer to those actions we take that can lead to consumers being better able to act for themselves to secure benefits from the market. This is all about consumers having the skills, confidence, and tools to better engage and benefit. Many of these sit well outside what a regulator can achieve. But some appear to be within our gift and within our remit.

The critical issues that might benefit from measures to strengthen consumer empowerment are:

- Making choices (information and tools to allow consumers to choose supplier and product);
- Exercising choices (switching); and
- Managing relationships with suppliers (e.g. knowing how to exercise rights).

Consumer protection and empowerment can impact on consumer outcomes, both through the effect on the individual consumer and through the impact on the effectiveness of the competitive process.

So a consumer who has been mis-sold a fixed-line telecommunications service will suffer harm directly through inconvenience, distress, and financial loss. But if the result of a prevalence of mis-selling is that the market acquires an unsavory reputation, then others who may benefit from switching may be put off from doing so, and they too will suffer harm as they will fail to benefit from a supplier

WE USE THE TERM “CONSUMER EMPOWERMENT” TO REFER TO THOSE ACTIONS WE TAKE THAT CAN LEAD TO CONSUMERS BEING BETTER ABLE TO ACT FOR THEMSELVES TO SECURE BENEFITS FROM THE MARKET.

or package that better suits their needs. And if, as a result of this bad reputation for the sector, consumers as a whole desist from switching, then all consumers will ultimately suffer as competitive pressure on suppliers is diminished and the ability of new entrants to disrupt the market is also reduced.

And the same is true for empowerment. A consumer who does not switch because of high informational, process, or contractual barriers to switching will potentially miss out on a superior deal. But if the market result is a lowering of competitive intensity due to reduced switching, then all consumers will eventually suffer.

III. What Issues Have Given Rise to Concerns in Our Recent Experience?

It is not always straightforward to perfectly classify issues that have given rise to harm into one category or another as sometimes consumer difficulties, and possible remedies, will span a number of categories. For example, we were faced with a particular issue of harm in our sector that came to be described as “mobile cash backs.” Consumers would purchase a mobile contract from an independent retailer. Because of complex incentives schemes, retailers were not able to undercut each other by lowering the network tariff, as this was not in their control. Instead, they competed by offering “cash back” to the consumer which, when done as a legitimate business practice, involves sharing with the consumer some of the commission that the network provider has paid to the retailer. However, over time the cash back promised became higher and sustainable only because some retailers had business models which involved actually paying the cash back only to a small number of consumers. One way they did this was to secure a low “redemption” rate for cash backs by putting terms in contracts that placed very onerous (and unreasonable) conditions on consumers to qualify for cash back—for example, requiring them to send the mobile bill to the retailer within an impossible timeframe.

Ofcom intervened to address this issue.¹¹ However, in our evaluation we considered both “consumer empowerment” and “consumer protection” remedies. Empowerment options included “educating” consumers as to the nature of the restrictive terms for receiving cash backs so that they could make informed decisions as to whether the restrictions justified the quantum of the cash back. Protection options involved prohibiting terms that did not appear reasonable.

We list below our main regulatory interventions in the telecoms industry over the last few years and whether we view them as mainly related to consumer protection or consumer empowerment.

A. CONSUMER PROTECTION

1. Luring Consumers Into Contracts or Services They Have Not Consented To

- **Fixed line mis-selling and slamming:** Consumers are mis-sold a service based on false information or aggressive and intimidating sales techniques, or are simply transferred to another supplier without their consent.
- **Mobile mis-selling:** Consumers are mis-sold a service based on false information or aggressive and intimidating sales tactics, or are misled into thinking that they are simply agreeing to an upgrade with their existing service provider when, in fact, they are being signed to a new contract with a different service provider.

2. Harm From Process Problems and Those That Occur Once Service is Being Provided

- **Silent calls:** An individual picks up the telephone and there is silence on the line. This problem is typically caused by features of automatic dialling equipment used by call centers—perhaps where the equipment has been configured irresponsibly. However, individuals have, in the past, been driven to believe that the silence is caused by something more sinister, such as a stalker or someone that is waiting to burgle the individual’s home.¹²
- **Mobile cash backs:** A consumer purchases a mobile contract from a retailer who promises to refund part of the line rental payable to the network provider during the life of the contract. But the consumer does not receive some or all of the cash back.¹³

3. Making it Hard for Consumers to Exit the Contract

- **Abuse of MAC process:** The broadband switching process often means the consumer needs a special code or a migration authorisation code (“MAC”) from their existing supplier that they must then give to their new supplier in order for the switch to take place. Abuse of

MAC process involves the existing supplier deliberately refusing to give the MAC out to prevent customers from leaving, or failing to invest in systems that allow for MACs to be given to customers.

- **Early termination charges:** These occur where customers leave within their contract period and are forced to pay an early termination charge. Customers sometimes are unaware that they are liable to such a charge, and sometimes the level of the charge is higher than could be considered to be “fair.”¹⁴

B. CONSUMER EMPOWERMENT

1. Making Choices

- **Broadband speeds:** Customers purchasing a broadband service advertised as e.g. “up to 8 mb/s” expect something close to that number but receive much less.¹⁵
- **Quality of service consumer information:** Customers find it difficult to make purchasing decisions as they have little reliable information on quality. Options that could assist to some degree include the publication of complaints to Ofcom on a provider-specific basis, market research, or publication of relevant provider data suitably audited to ensure reliability and comparability.
- **Accreditation of price comparison services:** Complexity of pricing has led consumers to be increasingly reliant on price comparison services (such as uSwitch and moneysupermarket.com). But these services make their money from commissions from those companies that consumers switch to, and this might create an incentive for the comparison service to present information that drives consumers to those that pay the highest commissions, rather than display accurate information presented with appropriate prominence.¹⁶

2. Exercising Choice

- **Strategic approach to consumer switching:** Existing product-specific processes for switching can be very difficult for consumers to navigate. The existence of bundles exacerbates the problems as product-specific processes can clash. But providers are often divided as to what new processes are appropriate for bundles given their different commercial incentives.
- **Rollover contracts:** Defined as a relatively new type of contract where consumers opt in to a fixed-term contract in return for a discount. With rollover contracts, at the end of the initial contract the customer is “rolled over” into a new fixed-term contract with an early termination charge. The result, if these contract types catch on, is that the market could feature a state where most consumers are locked into contracts that have only small windows in which they can change

providers. Behavioral economics plays an important role in our analysis. (See Section VI for a further description of these issues).

3. Managing Relationships With the Supplier

- **Complaints handling/ ADR:** Our sectors are complex and things will sometimes go wrong, often for technical or network reasons. However, consumers want to be with a provider that will be responsive when there is a problem and will take ownership of that problem. And, because consumers will find it expensive and intimidating to seek redress through the courts, the Communications Act allows us to require all providers to belong to an alternative resolution scheme approved by Ofcom which is free to the consumer and binding on the provider. However, if consumers are not made aware that they have this right, it will be of little value.¹⁷

IV. What Tools Do We Have to Improve Consumer Outcomes, Taking Account of the Impact of Market Mechanisms and the Role of Incentives?

A. WORKING WITH INDUSTRY

When competition appears not to be delivering effective outcomes for consumers in terms of information, switching, or protection, we will assess the extent to which we can work with industry to address the issues. This can involve a range of initiatives from reaching a shared understanding and objective with individual providers, to working with the industry to establish voluntary codes, or engaging the Office of the Telecoms Adjudicator (“OTA”) to develop and implement industry processes. The likely success of these routes can depend on the incentives for providers to address the problem and the credible threat of further intervention if progress is not made. Our work on co- and self-regulation has provided us with a toolkit to determine when the incentives for individual providers are likely to be sufficient to secure improvements through voluntary means.¹⁸

B. CONSUMER INFORMATION

We can sometimes use consumer information to address problems that emerge; for example, by alerting consumers to scams or explaining how to switch suppliers or make use of cooling-off periods. The information can take various forms, including items on our website, Ofcom consumer guides, information passed through consumer stakeholders such as Citizens Advice and Age Concern, and media articles.

We are also considering the extent to which comparative information to consumers might support the market by providing transparency and incentives to

address problems. Options here include publishing provider-specific research data and details of the number of complaints received by Ofcom about providers on particular issues.

In developing our approach in this area, we use behavioral economics to help us understand and test the role of such information and how consumers use information in markets.

C. INTRODUCING AND AMENDING "GENERAL CONDITIONS" ("GCs")

Proposing and introducing new GCs, or modifying existing GCs, allows Ofcom to set rules for communications providers and enables us to monitor and enforce those rules.¹⁹ This is a significant intervention. The Communications Act sets out tests for setting or modifying GCs, including requirements to ensure the change is objectively justifiable and proportionate. We have to focus very heavily on developing impact assessments of the costs and benefits of intervention. These can be challenging and time-consuming to establish. We are heavily reliant on providers to give estimates of costs, where they may well have an incentive to exaggerate estimates. Benefits, on the other hand, may be difficult to quantify.

This challenge may potentially be greater in our sector than in finance or energy, given relatively low spending by consumers on communications services and the complexity of networks and systems. Relatively modest interventions risk having a much larger impact on businesses (and therefore consumers), and solutions seem costly because of complex systems, with costs harder to absorb because of low average monthly consumer spending.

When considering options for new or amended GCs, we look at the scope for incorporating behavioral economics into our analysis of remedies and also for using experimental research techniques.²⁰ Potential areas where these approaches are particularly relevant include the quality of consumer information regarding service and broadband speeds; for example, where we consider how consumers actually acquire, absorb, process, and use information in their decision making rather than how they might have traditionally been assumed to respond as "rational consumers."

D. ENFORCING REGULATIONS

Although our powers under the Act in enforcing GCs are in theory considerable (with fines up to 10 percent of relevant turnover), in practice they are more limited. We must give operators the chance to remedy a breach or violation before issuing a fine, meaning that they could breach a GC but not face any sanction. As part of the Framework Review, the European Commission has proposed tougher enforcement powers for national regulators and these will need to be transposed into the relevant national legislation. The new powers would enable

us to fine an operator, even if they have subsequently remedied the breach. Tougher powers would provide stronger deterrents.

In addition to our powers under the Act to enforce GCs, we are empowered by the Enterprise Act 2002 (“EA”) to take enforcement action to stop infringements of certain consumer protection legislation, including the Unfair Terms in Consumer Contracts Regulations 1999 (“UTCCRs”), the Distance Selling Regulations 2000, and the Consumer Protection from Unfair Trading Regulations 2008 (“CPRs”).

Regulators have sometimes issued guidance on how they might interpret particular pieces of horizontal consumer law in particular sectors. The OFT has pioneered this approach with its various guidance on the UTCCRs, which has led to the high profile bank charges litigation.²¹ In 2008 Ofcom carried out a review of UTCCR guidance for our sectors. Entitled the “Additional Charges Review,” the review sought to give industry and consumers clarity on how we would interpret the UTCCRs when taking enforcement action on particular “additional” charges—those beyond the normal charge for the main service, such as early termination charges, fees for not paying by direct debit, or charges for paper bills.²²

Ofcom is obliged to consider complaints under the UTCCRs, so having guidance can be helpful to setting the boundaries of our likely actions. But while Ofcom can directly enforce breaches in the GCs, it can only enforce the UTCCRs (and similarly other consumer law) through the courts. Following the issuing of its guidance, Ofcom launched an enforcement and monitoring program. We have prioritized early termination charges with the result that fixed-voice providers have offered to make changes to their policies because of our intervention.²³

V. What Is the Evidence of the Effectiveness of Our Approach to Date?

Evaluating the effectiveness of our approach is not straightforward as there is rarely a single key performance indicator (“KPI”) that will measure the change in outcomes for consumers and link these outcomes causally to the actions we have taken. To address this, we seek to monitor a range of metrics and relate these, as best we can, to the actions that Ofcom has taken.

Each year, to facilitate accountability and discussion with stakeholders, we produce the Consumer Experience report as well as a number of research reports. Under a number of relevant headings, the Consumer Experience report contains all the metrics we have access to that help describe the consumer experience in our sectors. The reports are used to ensure we publicly evaluate our priorities going forward, that they are the right ones, and that the actions we are taking are well designed to achieve the desired outcomes. At the end of 2009 we also pub-

lished a Business Consumer Experience report²⁴ alongside the yearly Consumer Experience report.²⁵

VI. An Example—Analysis of British Telecom's Rollover Contracts

Many Communications Providers ("CPs") offer fixed-term contracts that require customers to commit to paying for a service for a minimum contract period ("MCP") in return for an incentive, such as an equipment subsidy—for example, a mobile handset or a set-top box—or a price discount. In order to exit these fixed-term contracts before the end of a MCP, customers usually have to pay an early termination charge.

Automatically Renewable Contracts ("ARCs") in communications markets are contracts where, at the end of a MCP (whether this is an initial or subsequent period), the contract rolls forward to a new MCP by default, unless customers proactively inform their CP that they do not wish this to happen.

ARCs are a feature in residential fixed-voice markets, and in business markets. Currently approximately 15 percent of U.K. residential fixed-voice consumers are on ARCs. British Telecom ("BT") has also introduced ARCs in the residential broadband sector. ARCs are not currently a feature of the mobile market. Since their introduction in the residential sector, Ofcom has had serious concerns about the potential harm that ARCs may cause, particularly if they become a widespread feature of this market.²⁶

We initially looked at ARCs in the context of our Review of Additional Charges published in December 2008.²⁷ The Review set out guidance on how we would enforce the UTCCR in the communications sector. Ofcom's current guidance states the conditions under which Ofcom believes ARC terms are more likely to be judged as "fair" under the UTCCR,²⁸ such as where the ARC term is transparent and a clear reminder is sent to the customer.

However, the test of fairness under the UTCCR is a legal test specific to those regulations and does not necessarily capture the full economic effects of a contract term. Consequently, our concerns about the effect of ARCs remained and we commissioned market research to better understand their effects and determine whether some form of intervention is appropriate.

Our initial market research, conducted in 2009, focussed on transparency and customer awareness in relation to BT's ARCs propositions, and included a mystery shopping exercise and a customer survey. This focus reflected the fact that, at the time, only a relatively small proportion of BT's ARC customers had rolled forward to a new contract (most contracts were sold in the second half of 2008) and the impact of ARCs on the switching process was not yet clear.²⁹

We are concerned that ARCs are damaging to consumers and competition in communications markets. We have identified two types of harm to consumers: a *direct effect* coming from the potential for ARCs to increase individual consumers' exposure to switching costs (in the form of an early termination charge) and an *indirect effect* coming from the potential for ARCs to lessen competition in the market, thereby reducing the pressure on firms to lower prices, and improve quality for all consumers.

While we recognise that ARCs may also have beneficial effects for some consumers—e.g. those who expect to remain with their supplier and who value the convenience of not having to renew their contract proactively—we believe these benefits are relatively limited and are outweighed by the costs.

WE ARE CONCERNED THAT ARCS
ARE DAMAGING TO CONSUMERS
AND COMPETITION IN
COMMUNICATIONS MARKETS.

Towards the end of 2009 Ofcom commissioned Professor Gregory S. Crawford and ESMT Competition Analysis to conduct an econometric analysis of BT customer data in order to identify whether BT's ARC term had an impact on customer switching. An econometric approach was necessary in order to isolate the impact of the ARC term itself, as separate from other factors such as the price discount associated with the offer, and changes in the competitive dynamics in the market.³⁰

The econometric analysis indicated a clear causal link between ARCs and reduced levels of consumer switching. Furthermore, it showed that the effect was separate from the impact on switching levels of other factors such as price discounts. We believe this effect stems from the opt-out nature of the process for contract renewal, rather than a lack of transparency surrounding ARC terms or the complexity of the process for opting out. Because it stems from such a core aspect of ARCs, this indicates that any example of such a contract is likely to be harmful to consumers and to effective competition.

We therefore proposed in our March 2011 consultation an amendment to General Condition 9 that will prohibit “opt-out” processes for MCP renewal (processes where end users automatically enter a new MCP by default unless they proactively inform their CP that they do not wish this to happen) in any form in the fixed voice and broadband sectors.³¹ ▼

1 OFT, *What does Behavioural Economics mean for Competition Policy*, available at http://www.of.gov.uk/shared_of/economic_research/oft1224.pdf.

2 John Fingleton, *Joining Up Competition and Consumer Policy. The OFT's Approach to Building an Integrated Agency*, available at http://www.of.gov.uk/shared_of/speeches/2009/spe-1209.pdf. This has also recently been discussed by Eliana Garces, *The Impact of Behavioral Economics on Consumer and Competition Policies*, 6(1) COMPETITION POL'Y INT'L, pp. 145-152 (Spring 2010) and Michael Salinger, *Behavioral Economics, Consumer Protection, and Antitrust*, 6(1)COMPETITION POL'Y INT'L, 65-86 (Spring 2010).

- 3 See, for instance, Mark Armstrong, *Interactions between Competition and Consumer Policy*, OFT Discussion paper available at http://www.ofcom.gov.uk/shared_ofcom/economic_research/ofcom991.pdf or Stefano Della Vigna, *Psychology and Economics: Evidence from the Field*, (47) J. ECON. LIT, pp. 315-372 (June 2009).
- 4 See, for instance, the recent Mountain View Learning and OFT study, *Pricing Practices: Their Effects on Consumer Behaviour and Welfare*, available at http://www.ofcom.gov.uk/shared_ofcom/business_leaflets/659703/Advertising-of-prices/Pricing-Practices.pdf.
- 5 Ofgem has recently proposed to require electricity and gas providers to offer at least one simple standard tariff to enable consumers to compare the competing offers, see their recent proposals for the retail markets at <http://www.ofgem.gov.uk/Markets/RetMkts/rmr/Pages/rmr.aspx>.
- 6 Behavioral economics suggests consumers have particular difficulty making forward-looking assessments. Grubb finds evidence that mobile customers are overconfident about not exceeding usage allowances, and choose wrong tariffs as a result. See Grubb, *Selling to Overconfident Consumers*, 99(5) AMER. ECON. REV, pp. 1770-1807 (2009).
- 7 See, for instance, the discussion in Section 2 of Ofcom's Strategic Review of Consumer Switching, available at <http://stakeholders.ofcom.org.uk/binaries/consultations/consumer-switching/summary/switching.pdf>.
- 8 See, for instance OECD (2011), *Broadband Bundling: Trends and Policy Implications*, OECD DIGITAL ECONOMY PAPERS, No. 175, available at <http://www.oecd-ilibrary.org/docserver/download/fulltext/5kghtc8znnbx.pdf?expires=1301393287&id=0000&accname=guest&checksum=E2C8B1987461-BA79B14CC61404CA0101>.
- 9 Section 3(4) of the Communications Act 2003 states that, in performing its duties, Ofcom must also have regard to a number of matters as appears to be relevant in the circumstances; in particular, the desirability of promoting competition in relevant markets.
- 10 Ofcom Statement, *Tackling Abandoned and Silent Calls*, (October 1, 2010), available at <http://stakeholders.ofcom.org.uk/binaries/consultations/silentcalls/statement/silentcalls.pdf>.
- 11 Ofcom Statement, *Protecting consumers from mis-selling of mobile telecommunications services*, (March 17, 2009), available at <http://stakeholders.ofcom.org.uk/binaries/consultations/mobmisselling/statement/statement.pdf>.
- 12 *Supra* note 10.
- 13 *Supra* note 11.
- 14 The main UK landline providers significantly reduced their early termination charges following discussions with Ofcom in 2010, (<http://consumers.ofcom.org.uk/2010/06/cheaper-charges-for-uk-consumers-to-end-phone-contracts/>).
- 15 Ofcom has published research into the broadband speeds actually offered by fixed broadband providers. See a recent example at <http://media.ofcom.org.uk/2011/03/02/average-broadband-speed-is-still-less-than-half-advertised-speed/>.
- 16 See a summary of Ofcom's accreditation work at <http://stakeholders.ofcom.org.uk/consultations/ocp/statement/pricescheme/consumerfaq/>.
- 17 Ofcom is currently reviewing the effectiveness of alternative dispute resolution schemes. See the recent call for inputs at <http://stakeholders.ofcom.org.uk/binaries/consultations/alternative-dispute-resolution/summary/adr.pdf>.

- 18 Ofcom Statement, *Identifying appropriate regulatory solutions: principles for analysing self- and co-regulation*, (December 2008), available at <http://stakeholders.ofcom.org.uk/binaries/consultations/coregulation/statement/statement.pdf>.
- 19 When the new EU communications regime was implemented in the United Kingdom on July 25, 2003, individual licenses granted under the Telecommunications Act 1984 were replaced by the General Authorization regime. The effect was that licenses were no longer required for providing communications networks or services in the United Kingdom—everyone is “generally authorized” to do so. However, the General Authorization is subject to the General Conditions of Entitlement; these conditions apply to all persons providing electronic communications networks and services.
- 20 Ofcom has published two laboratory experiments on consumer behavior, see <http://stakeholders.ofcom.org.uk/binaries/consultations/consumer-switching/annexes/economics-research.pdf> and <http://stakeholders.ofcom.org.uk/market-data-research/telecoms-research/experiments/>. See also the more general discussion at <http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/experiments.pdf>. We are also currently carrying out additional experiments that we plan to publish in the next few months.
- 21 OFT, *Unfair contract terms guidance*, (September 2008), available at http://www.oft.gov.uk/shared_of/reports/unfair_contract_terms/of311.pdf.
- 22 Ofcom Statement, *Ofcom review of additional charges*, (December 2008), available at <http://stakeholders.ofcom.org.uk/binaries/consultations/addcharges/statement/addchargestatement.pdf>.
- 23 *Supra* note 11.
- 24 Available at <http://stakeholders.ofcom.org.uk/market-data-research/market-data/consumer-experience-reports/bce/>.
- 25 Available at <http://stakeholders.ofcom.org.uk/market-data-research/market-data/consumer-experience-reports/consumer-experience/>
- 26 Ofcom’s consultation on its proposal to ban ARCs is available at <http://stakeholders.ofcom.org.uk/binaries/consultations/arcs/summary/arcs.pdf>.
- 27 See <http://stakeholders.ofcom.org.uk/binaries/consultations/addcharges/statement/addchargestatement.pdf>
- 28 See ¶ 97 onwards in Ofcom’s Guidance, available at <http://stakeholders.ofcom.org.uk/binaries/consultations/addcharges/statement/Guidance.pdf>.
- 29 See Annexes 11, 12, and 13 at <http://stakeholders.ofcom.org.uk/consultations/arcs/>.
- 30 See the report at http://stakeholders.ofcom.org.uk/binaries/consultations/arcs/annexes/Annex_07.pdf.
- 31 The OFT has also recently published a study on consumer contracts where the effect of rollover terms is discussed, see <http://www.oft.gov.uk/OFTwork/markets-work/current/consumer-contracts>.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Applications Want to be Free: Privacy Against Information

Michael R. Hammock
Middle Tennessee State University
&

Paul H. Rubin
Emory University

Applications Want to be Free: Privacy Against Information

*Michael R. Hammock & Paul H. Rubin**

The debate over online privacy pays too little attention to the costs and benefits of the current systems of privacy protection and advertising-supported online applications. The costs of online privacy-related harm (such as identity theft) and of protective activities are small relative to the benefits from applications that are supported by online advertising, which depends on the collection of personal information. Advocates of increased privacy focus too much on increased privacy as a solution, and not enough on alternative forms of information security. Surveys show that consumers do not like targeted advertising, or the information collection that allows it, but this may be a form of rational irrationality. That is, it may not pay for consumers to understand the costs and benefits of reduced information use.

*Michael R. Hammock is an Adjunct Professor of Economics at Middle Tennessee State University; Paul Rubin is the Samuel Candler Dobbs Professor of Economics at Emory University

I. Introduction

Both Europe¹ and the United States² are considering regulation that would increase consumer privacy and make the collection of personal information more difficult. Therefore it is worth examining whether these regulatory changes make economic sense.

Privacy advocates have pointed to identity theft as a reason to increase online privacy. They propose to make it illegal to collect information about consumers, by mandating opt-in as a default rule, or by other regulatory changes. These suggested policy changes seem not to be based on economic theory or on evidence beyond anecdotes. In this paper, we propose and defend the following assertions:

1. Proponents of increased privacy have not made a case based in economic theory or evidence, are vague regarding harm caused to consumers by lost privacy, and sometimes demonstrate fundamental misunderstandings of basic economics and the relevance of information security.
2. The total benefits of the current “opt-out” default rule (which requires consumers to take action to prevent the collection of their personal information) exceed the total costs, although it is not possible to tell if the marginal benefit of increased privacy equals the marginal costs. While some alternative approaches (such as a complete prohibition on information collection) to protecting personal information are undoubtedly inefficient, for some others (such as “quid pro quo”) the data and theory do not allow us to make a prediction as to their efficiency. Nonetheless the fact that the collection of personal information has generated such a huge surplus of benefits in excess of costs suggests that we should be reluctant to impose fundamental changes.
3. Surveys of consumers suggest that consumers dislike both targeted advertising and the information collection that allows it. We contend that these surveys may have problems and that, even if the surveys are correct, consumers may be displaying “rational irrationality.” Their opinions on privacy regulation may be no more reasonable than their opinions on international trade. If consumers do have valid privacy concerns, markets can and do respond to them.

II. The Arguments of the Privacy Advocates

We have not been able to find any privacy advocates making sensible economic arguments for increased privacy. As far as we can tell, arguments for increased online privacy are based on rights (rather than efficiency) and anecdotes (rather than data). Walker³ also complained of a lack of cost-benefit analysis in discussions of privacy rights, and Szoka & Thierer⁴ point out that the harm privacy advocates worry about is conjectural or speculative, rather than concrete. Lenard & Rubin⁵ provide an overview of how information collection and targeted adver-

tising work, and argue that the benefits of more relevant ads are large, while the costs are small.

Hahn & Layne-Farrar⁶ divided the participants in the online privacy debate into four positions. The distinctions between some of these categories are fuzzy; we simplify them into two: The “Increased Privacy” camp and the “Status Quo” camp. The Increased Privacy camp wants to make opt-in the default rule and wants to limit the use of data to the task for which it was originally collected. This would mean that consumer information cannot be collected without the consumer explicitly choosing to allow it. Also data could not be used for any task other than the task immediately at hand, and could not be resold or reused without explicit permission from the person described by that personal information. We will call the other side of this argument the “Status Quo” camp. The Status Quo camp argues that the benefits of

information collection under the current system exceed the costs, and that market responses will take care of any problems.

WE HAVE NOT BEEN ABLE TO FIND
ANY PRIVACY ADVOCATES MAKING
SENSIBLE ECONOMIC ARGUMENTS
FOR INCREASED PRIVACY.

When members of the Increased Privacy camp argue for restrictions on the re-use of personal information, or for a switch to “opt-in” as a default rule, their arguments are generally based on an implicit right on the part of consumers not to have any information collected about them without their knowledge and consent. This is reflected in the European view of privacy regulation as well. As *The Economist* put it, “European regulations are inspired by the conviction that data privacy is a fundamental human right and that individuals should be in control of how their data are used.” Regarding U.S. regulation, when Marc Rotenberg of the Electronic Privacy Information Center (“EPIC”) testified on December 2, 2010 to the Committee on Energy and Commerce, his argument seemed to be that firms collect a *lot* of information, and consumers don’t know this.⁷ In 2008 he argued that “the detailed profiling of Internet users violates the fundamental rights of individuals, diminishes the accountability of large corporations, and threatens the operation of democratic governments.”⁸ There is no discussion of benefits and costs—are consumers genuinely being harmed? Do they benefit in any way?

At a Federal Trade Commission (“FTC”) roundtable discussion, a panelist from the Privacy Rights Clearinghouse shared a few examples of horrifying cases in which information collected online was used for criminal means, including stalking and rape.⁹ As terrible as these cases may be, however, anecdotes are less persuasive than data. Basing policy on anecdotes will result in a bias toward regulating; millions of people uneventfully going about their business online do not make for interesting counter-anecdotes. Furthermore, the data suggest that online identity fraud is rare. We will return to the costs of identity fraud shortly.

Gellman¹⁰ argued that consumers have revealed their preference for privacy through their willingness to pay for it, in the form of unlisted numbers, caller ID, spam filters, and sorting through junk mail. Some of these costs are dated now,

with do-not-call lists to stop telemarketers, the ubiquity of cell phones (with caller ID built in), and very effective automatic spam filters for free email accounts such as Gmail—technology and policy have already caught up with many of these problems. Gellman then adds up the costs of pursuing extreme privacy—anonimization service, identity theft protection, reports from all three credit bureaus, credit monitoring, and so on. The total annual costs for a single consumer are nearly \$300, but what are we to make of this? The costs of pursuing such extreme privacy are very high, but this is like arguing that the roads are not safe enough by citing the high cost of an armored car. Consumers who do not incur these costs face an extremely small chance of identity theft occurring, as we discuss later. Consumers are wise to forego all these expenses unless they put an extremely high value on safety.

Privacy advocates are not always so explicit about the costs of lost privacy. In the Center for Digital Democracy’s comments submitted to the FTC regarding privacy regulation,¹¹ the word “cost” appears six times, yet in none of those cases are the costs of lost privacy described or explained. Rather, it is asserted that researchers who attempt to determine the costs and benefits of behavioral advertising (which depends on the collection of personal information) “misunderstand” the costs for consumers—without explaining how they have erred. The authors seem to suggest that the fact that targeted ads can now be targeted accurately and delivered very quickly is itself cause for action.

What of the costs to consumers of information breaches? In 2009 Mark Rotenberg of EPIC testified to the House Commerce Committee¹² regarding legislation that would regulate notification of data security breaches. Rather than focusing on information security, however, Professor Rotenberg also talked about making it more difficult for corporations to collect and use personal information in the first place. It is true that preventing corporations from collecting personal information (or preventing it from being in their interests to collect personal information) would reduce the damage from data security breaches. This is like arguing that doing away with privately owned cars would be a means to reduce automobile accidents—the cure would be worse than the disease. Whether data security should be regulated differently is a good question, and one that EPIC has addressed in the past (as with the case of TJX).¹³

A growing body of economic literature examines information security. Anderson & Moore¹⁴ provide a good overview of the fundamental economic issues. The core problem is that people with the responsibility to protect data may not face the full costs of failing to do so—there may be a negative externality, resulting in inefficiently lax security. It is not clear that this is the case; Lenard and Rubin¹⁵ argue that the costs of breach fall almost entirely on firms

THE CORE PROBLEM IS THAT PEOPLE WITH THE RESPONSIBILITY TO PROTECT DATA MAY NOT FACE THE FULL COSTS OF FAILING TO DO SO—THERE MAY BE A NEGATIVE EXTERNALITY, RESULTING IN INEFFICIENTLY LAX SECURITY.

that store information. This could mean that investment in security is slightly suboptimal, unless the externality is inframarginal, in which case it need not be suboptimal at all. Even if there is a market failure, it seems more reasonable to address this market failure than to shrink the market for personal information. To put it succinctly, if members of the Increased Privacy camp are concerned about data breach, the lowest cost way to address this is to improve the incentives to provide security rather than to limit the collection and use of information. How could this be accomplished?

There are several possible regulatory tools available to improve information security. For example, altered breach disclosure laws could allow both consumers and firms to be both more proactive as well as react more quickly, although Romanosky et al.¹⁶ find that state-level variation in breach disclosure laws have only a small effect on identity theft. Anderson et al.¹⁷ suggest mandated disclosure of vulnerabilities. Firms suffering breach could be assigned liability for all damages caused, leading them to internalize the security externality.¹⁸ In the case of TJX, EPIC suggested the assignment of \$10 million in additional civil penalties, but it is not clear why liability for damages caused would not be the efficient remedy.¹⁹ Large breaches are hard to hide, so it is not as though high damages were necessary to maintain efficient expectation damages.

There are a variety of other options available as well, but our purpose is not to catalog them or to assess them, but to point out that if information security is the problem, the debate should center on the means to address this problem. Privacy concerns should not distract from the debate over what regulatory tools to use, if any, to improve information security, and reduce the costs of breach. Privacy advocates have missed or ignored this point.

Members of the Increased Privacy movement further neglect the *benefits* created by the current system of information collection, which supports personal ads, which in turn support free online applications. At the heart of economics is the idea that *incentives matter*, and if the money that funds online applications is reduced, or if the returns from developing these applications are reduced, fewer

MEMBERS OF THE INCREASED
PRIVACY MOVEMENT FURTHER
NEGLECT THE *BENEFITS*.

online applications will be provided. Privacy advocates do not seem to see the connection, and when confronted with it, deny it. For example, when interviewed by ABC News, privacy and security advocate Christopher Soghoian asserted “The web was free for the last 15 years

before they were tracking people, and it will continue to be free after they track people.”²⁰ The web that was free in the 1990s was very different from the web today, with its wide variety of online applications. Advocates act as if the question is “either/or” (will there be an internet or not?) when it is actually “How much?” (what sort of functions will the internet perform?). Economists should not assume that everyone understands supply curves slope upward. Again, we will return to the scope of the benefits of these applications shortly.

There has also been concern about sites like Spokeo.com.²¹ Spokeo collects personal information from a variety of online sources, including social networking sites. It purports to have information on income, wealth, property value, number and age of people in the household, addresses, phone numbers, email addresses, and other personal information. For users to access any information beyond the basics requires paying a fee. This causes concern because of fears of identity fraud, and perhaps a general “creepiness” from finding out that other people can obtain information about oneself. It is important to keep in mind, however, that it has long been possible to pay someone to find information about other people. This is not a new phenomenon; it is an old phenomenon that has moved to a new medium.. Finally, we should keep in mind that this information contained online was either put there willingly by consumers themselves, or it is public record (such as property records).

What can we say about the balance between the Privacy Advocates and the Status Quo proponents? What are the costs and benefits to consumers of lost online privacy? Consumers are clearly harmed by online identity fraud, which occurs when someone is able to impersonate the consumer, gaining access to his or her accounts. In addition to creating debts for the consumer, the consumer’s credit record may be harmed, and resolving these problems may create additional expense. In the next section, we discuss the size of these costs, as well as the benefits of the current system.

III. Opt-In, Opt-Out, and the Costs and Benefits of Targeted Advertisements

The current system of privacy protection in the United States is “opt-out.” Consumers must take actions to prevent personal information from being collected, by: running software or establishing non-default settings that routinely remove cookies; explicitly telling websites not to use their data (when such an option is available); refraining from putting personal information on sites like Facebook; and taking whatever other measures they can. This is a default rule. That is, by default, consumers are assumed to have given permission to collect personal information; they must intentionally opt out to deny permission.

Advocates of increased privacy argue that an “opt-in” default rule would be superior. Consumers would have to give explicit permission any time their information was collected, sold, used, or reused. Groups such as the Consumer Electronics Association, Consumer Watchdog,²² Center for Democracy and Technology,²³ and the Center for Digital Democracy and U.S. Public Research Interest Groups²⁴ favor opt-in as the default rule. The 2009 changes to the European Union’s E-Privacy Directive²⁵ require that cookies should only be stored on a user’s computer if the user consents, and Europe’s privacy regulation

prior to that was generally more restrictive of the collection and use of personal information than that in the United States.

If transactions costs were zero, the decision as to which default rule to apply would not matter; Coasean bargaining²⁶ would put the personal information in the hands of firms if the firms valued the data more highly than consumers valued their privacy. The transactions costs are not zero, however. Under both opt-in and opt-out consumers must take the time to become informed, make a decision, and implement it across all their computers and software. If the transactions costs are too high, consumers and firms will not be able to bargain to the efficient outcome. It is therefore important that the rights be assigned efficiently by law and regulation. What, then, is the efficient default rule: opt-out, or opt-in? Do the benefits of information collection under the current opt-out default rule exceed the costs? What would happen under opt-in?

Bouckaert & Degryse²⁷ develop a theoretical model that suggests that opt-out is the efficient default rule unless the costs of opt-in are zero (in which case opt-out and opt-in are equally efficient). This is because fewer consumers buy from the socially optimal supplier under opt-in, and they pay higher prices as a result.

There are several empirical studies of opt-in and opt-out. Staten and Cate (2003)²⁸ conducted a case study of MBNA (a bank subsequently bought by Bank of America), finding that opt-in would make it more difficult to match credit offers to customers, and make it more difficult for MBNA to fight fraud. The authors did not examine the effect on consumers, but two outcomes are likely: consumers could receive more credit offers of a less targeted (and therefore less useful and more annoying) nature, and fewer consumers would get the credit appropriate for their personal needs.

Johnson & Goldstein²⁹ found that there is a 16 percent increase in organ donations in countries in which opt-out (that is, one must take action to prevent oneself from being an organ donor) is the default rule, relative to countries in which opt-in (one must take action to become an organ donor) is the default rule. This is despite the fact that opting in or opting out is often no more costly than checking a box on a driver's license application. The simple switch of the default rule can be more effective than campaigns to encourage people to opt into donating. Thaler & Sunstein³⁰ found a similar result for 401 (k) plans: Enrollment increases dramatically when moving from opt-in to opt-out. Clearly default rules matter; people may not opt-in simply to avoid the costs of having to think about it. In the case of organ donation, this means people avoid thinking about death. With online privacy, it means that consumers avoid thinking about the costs and benefits of allowing personal information to be collected and, perhaps more importantly, they avoid thinking about the very small chance that their data might in some way be abused. The possibility that poorly chosen default rules can allow consumers to *avoid* making careful decisions does not encourage us to believe that the outcome of opt-in will be efficient.

Given this “sticking” of default rules, it is important to choose the right one. If the value of privacy is greater than the value of targeted advertising and the online applications that targeted advertising funds, then opt-in is the efficient default rule. If the value of targeted advertising and online applications is greater than the value of privacy, then opt-out is the efficient default rule. There are several empirical studies that can help determine which is the case.

First, what are the measureable, concrete costs to consumers of online breaches³¹ and identity fraud? The 2010 Javelin Identity Fraud Survey Report³² found that damage from identity fraud (both online and offline) was \$54 billion in 2009. For the sake of comparison, a 2003 FTC report³³ found that the total costs from identity fraud were \$52.6 billion, of which \$5 billion was losses to consumers, and \$47.6 billion was losses to business. A 2006 report³⁴ found that the losses were only \$15.6 billion total, but survey methods changed, and costs were not broken down by incidence.

We will assume the high cost estimate of \$54 billion from the 2010 Javelin report. The 2010 Javelin report preview does not provide the fraction of cases in which personal information was obtained online, but the 2008 report³⁵ says that 12 percent of identity fraud in 2008 was accomplished by information obtained online. This number comes from victims who knew how their personal information was obtained; it may be the case that victims who do not know how their information was obtained were more or less likely to have had it taken online, but we will use the 12% number as it is the best we have. This means that around \$6.48 billion in damage from online identity fraud was inflicted in 2009. Compared to the costs of fraud overall, the size of the online economy,³⁶ or the overall economy, this is not an enormous cost. With around 220 million Americans online,³⁷ that works out to about \$29.44 in online identity fraud damage per user.

A 2010 IAB Europe study³⁸ found that the value to consumers of preventing online ad disturbance (defined as the risk of abusing personal information and the annoyance of advertisement intrusion) is around EUR 20 billion, or around \$28 billion. It appears that this number is the sum of the value of protection in the United States and Europe, and separate numbers are not provided. For sake of argument, and to be conservative, let us assume that the entire \$28 billion applies to the United States alone—that is, for U.S. consumers, the value of avoiding the costs of having their information collected is around \$28 billion.

What are the benefits to consumers of advertising-funded applications online? The IAB report finds that, after netting out the costs of disturbances and paid services (including internet access), consumer surplus from web services is around \$100 billion for the United States and Europe combined. More than half of this consumer surplus comes from free services. Again, they do not provide separate consumer surplus estimates for Europe and the United States,³⁹ although

they do show that there are differences across countries in the fraction of consumer surplus generated by different online services. The report also projects that consumer surplus will continue to grow at around 13 percent per year, based on current trends.

This shows that the consumer surplus of the current regimes in the United States and Europe have enormous benefits in excess of costs, but what of the difference between U.S. and European privacy policies? Goldfarb & Tucker⁴⁰ find that European privacy regulation reduces the effectiveness of targeted online advertising, resulting in ads that are less relevant to consumers and generate less revenue. This reduced effectiveness may also result in more ads being served; in order to raise consumer purchase intent by the same amount as an ad prior to the tightening of E.U. privacy regulation, an advertiser must buy 2.85 times more advertising. Goldfarb & Tucker estimate that by changing the privacy regula-

SECOND, THE COSTS OF
IDENTITY FRAUD COMMITTED
ONLINE—A CONCRETE,
MEASUREABLE PRIVACY CONCERN—
APPEAR TO BE RELATIVELY SMALL.

tions in the United States, revenue from online advertising could fall from \$8 billion to \$2.8 billion. If ads become less effective, and generate less revenue, then we should expect less funding for ad-supported applications, and a loss of value to consumers.

We draw several conclusions from this body of research. First, a switch to opt-in as a default rule would likely result in a dramatic reduction in the amount of information collected, and this would cause targeted ads to be less valuable. Second, the costs of identity fraud committed online—a concrete, measureable privacy concern—appear to be relatively small. Third, the benefits to consumers of online services such as search, free email, Google docs, mapping services, Facebook, search, and so on, are enormous. Decreased advertising revenue would reduce the incentive to provide these online services or reduce their quality.

There is an important caveat, however. Some of these online applications might persist without targeted advertising. We know that the total benefits of the current system exceed the total costs, but we cannot be sure that the marginal benefits equal the marginal cost. Currently Europe's privacy regulations, though stricter than in the United States, are not *radically* stricter. We do not have the data to tell us whether a marginal change toward slightly more privacy creates benefits greater than costs. Radical changes are more likely to reduce the benefits of free online applications (supported by targeted advertising). Still, we cannot be sure what sort of equilibrium would emerge if a radically different system, under which consumers were paid for their personal information (perhaps with access to online applications), were implemented. However, since the current system evolved in a free market situation, it is unlikely that any radically different alternative would be preferable.

IV. Consumer Views of Privacy and Targeted Advertisements

Surveys often show that consumers do not want advertisement targeted toward them, and that they do not feel their loss of privacy is worth any of the benefits provided. A 2008 Harris Interactive/Westin Survey⁴¹ asked survey respondents how comfortable they were with sites like Google using personal information to tailor advertisements to their interests, using the revenue to provide free services like email. 59 percent of consumers were not comfortable with this. When asked if they would be comfortable with targeted ads if a list of privacy protections were implemented, 55 percent said they would be comfortable.

A December, 2010 Gallup Poll⁴² of U.S. internet users found similar results: 67 percent opposed targeted ads based on behavioral tracking, and 61 percent did not believe that the support for free online services made possible by targeted ads justified their use. 61 percent of users reported having seen such ads, and 90 percent of them stated that they paid little or no attention to them. Strangely, a plurality of users said they would prefer to allow advertisers of their choice to target them, as opposed to allowing all advertisers or no advertisers.

Turow et al.⁴³ conducted a survey and found that 66 percent of respondents did not want websites to show them ads tailored to their interests, although 47 percent would like sites to give them discounts tailored to their interests.

Consumers were more accepting of ads that were targeted based on the site they were currently visiting, but not of ads based on sites they had previously visited. Younger respondents were more accepting of targeted ads, but still had 55 percent opposition. Survey respondents were also strongly in favor of laws increasing their online privacy. They did not understand current regulations, however, believing that the law provided more privacy than it actually does. For example, 54 percent believed incorrectly that websites with privacy policies must delete one's personal information if one asks them to do so.

Spiekerman et al. (2005)⁴⁴ surveyed consumers in 2000 about their privacy preferences and their behavior, using an online shopping experiment. They found that while most consumers expressed privacy concerns, their behavior did not “live up to their self-reported privacy preferences.” They provided personal information for no clear reason—even some users categorized as privacy fundamentalists. Aquisti & Grossklags⁴⁵ conducted a survey and found that 87.5 percent of consumers who said they were highly concerned about the collection of personally identifying information (like a name or address) signed up for a shopping loyalty card—which required using their real personal information. Of those respondents concerned about credit card and identity fraud, only 25.9 per-

STRANGELY, A PLURALITY OF USERS SAID THEY WOULD PREFER TO ALLOW ADVERTISERS OF THEIR CHOICE TO TARGET THEM, AS OPPOSED TO ALLOWING ALL ADVERTISERS OR NO ADVERTISERS.

cent used credit alert features. Of those who said that consumers should use technology to protect their privacy, 62.5 percent said they had never used encryption⁴⁶ and half never used shredders to destroy documents containing personal information. Clearly there is a disconnect between what consumers say and do.

McDonald & Cranor⁴⁷ conducted in-depth interviews with 14 subjects regarding internet advertising and privacy. They found that the consumers disagreed over what constitutes an advertisement, and sometimes do not recognize ads for what they are. They do not understand exactly what cookies are, how they work, how information about their browsing behavior is collected, and only three of them understood that cookies were related to targeted advertisements. Some subjects preferred ads that were more relevant, while others were concerned about the privacy implications of targeted ads. Regarding specific harms of lost

privacy, users identified the loss of privacy itself as the primary harm, with one user suggesting concern over privacy would cause users to withdraw from online life.

CONSUMER OPINION, WHILE
CERTAINLY IMPORTANT FOR
POLICYMAKERS (PARTICULARLY
THOSE LOOKING FOR VOTES),
IS NOT NECESSARILY A GUIDE
TO EFFICIENT POLICY.

Members of the Increased Privacy movement quite reasonably cite these robust survey results as an argument for stricter regulation. We believe that this position is incorrect, however.

Consumer opinion, while certainly important for policymakers (particularly those looking for votes), is not necessarily a guide to efficient policy. As Bryan Caplan⁴⁸ has shown, consumers-as-voters are often *rationally irrational*; they often support policies that make little economic sense, such as agricultural subsidies, and disagree with experts (economists, toxicologists, climatologists, etc.) despite lacking the information on which to base an informed opinion.

A better phrase to describe this phenomenon would be *rational systematic bias*.⁴⁹ Consider free trade, for example. Most economists favor free trade, and believe that the benefits of reducing trade barriers outweigh the costs. They base this on hundreds of years of theory and evidence. If poorly informed laypeople were *rationally ignorant*, then we would expect some of them to think that free trade is less beneficial than it actually is, while an equal number would think that free trade is *more* beneficial than it actually is. This is not what we observe, however. Voters' views on trade are *systematically* biased; they err consistently on the side of believing that trade is *bad*. Averaging the opinions of all the voters does not result in something close to the truth; it results in an average opinion that is biased away from the truth (with truth, in Caplan's model, being represented by the averaged opinions of experts).

This is rational, Caplan argues, because voters do not face the cost of holding incorrect beliefs. Their one vote will not change policy, and when it comes to policy issues, holding unpopular (but more correct) opinions will not benefit

them. Voters therefore hold (incorrect) opinions as a result of culture or the early evolutionary environment.⁵⁰ Consumers making shopping decisions are faced with a very different situation: they face all the costs and benefits of their decisions. We expect them to be better informed, because holding incorrect beliefs is costly. If a product is risky, consumers take action to protect themselves, such as paying a home inspector to make sure the house they are looking at has no hidden dangers, or hiring a mechanic to make sure the car they are about to buy is fully functional. Consumers collect product information and reviews to help make decisions while shopping online. They do these things because the costs of poor decisions, and the benefits of good decisions, fall entirely on them.

How, then, does this relate to online privacy? It is natural for consumers to be uncomfortable with the idea that someone is collecting information about them. We are not used to the idea of a machine collecting data, which is then fed through algorithms and used, impersonally, to send us advertisements. Consumers' reaction is concern, and they support policy changes to increase their privacy. In two books,⁵¹ Clifford Nass has carefully shown that people fundamentally misunderstand the nature of intelligent machines. For example, people are more likely to rate a computer's performance as good if they are asked while working on that specific computer than if they are asked while working on a different computer.⁵² That is, people are "polite" to computers. We hypothesize that the same principle applies to tracking by websites: people cannot conceive of being tracked by a machine, and instead respond as if some human knows what they are doing. Our brains did not evolve to understand the nature of relatively intelligent machines, and we treat them as if they were people.

This instinct does not necessarily make for good policy, however. The available data on costs and benefits suggest that the risks of having data on one's browsing habits collected are low; the damage from identity fraud is relatively small. It is hard to believe that consumers recognize the extent to which free online sites and applications are funded by advertising. There is a free rider problem here, as well. When asked individually, a consumer might prefer not to be tracked, and thereby obtain a free ride off of the creation of online applications funded by advertising targeted at *other* consumers.

Public opposition to the online collection of personal information is not *per se* evidence that consumers are being harmed and need regulation to protect them, just as voter support for agricultural subsidies is not evidence that we would run out of food without such subsidies. To put it another way, surveys have shown that consumers do not understand how cookies and online information collection techniques work. They have also shown that consumers see information col-

OUR BRAINS DID NOT EVOLVE TO UNDERSTAND THE NATURE OF RELATIVELY INTELLIGENT MACHINES, AND WE TREAT THEM AS IF THEY WERE PEOPLE.

lection as dangerous. Why do privacy advocates consider the second result to be evidence of consumer wisdom, given the first result?⁵³ Would it not be more reasonable to conclude that consumers' views of information collection are of dubious value?⁵⁴

This raises another point. Privacy advocates often claim that if consumers fully understood how much information was collected and how it was being used, then they would be much more concerned. But the very ignorance of consumers is itself evidence of the lack of harm. Consumers learn about things that are actually harmful, such as tainted foods or dangerous products. The fact that consumers do not bother to learn about data collection is itself evidence that this process is not harmful. Privacy advocates have for many years been warning consumers about this danger, but consumers have blithely been ignoring these warnings, because they have not observed or suffered any real harm.

BUT THE VERY IGNORANCE
OF CONSUMERS IS ITSELF EVIDENCE
OF THE LACK OF HARM.

If consumers desire greater online privacy, entrepreneurs should find it rewarding to provide protective services. In fact, there are a variety of tools available to consumers right now. Based on our own casual experience, we have noted that Google Chrome has an incognito mode, which does not either record webpages or files downloaded in browsing or maintain download histories, and deletes cookies after the window is closed. Firefox has a similar Private Browsing mode, and Internet Explorer 8 has an InPrivate Browsing mode. The Dolphin browser for Android devices has an option to delete cookies automatically after each session.

These modes do not prevent all tracking, but they can drastically reduce the amount of information collected, at a very low cost (an occasional extra click, at most). Future versions of Firefox and Internet Explorer will support the Do Not Track flag, although this does not work unless websites support it, and whether they will do so remains to be seen. People can easily add a free Gmail or Yahoo email account and use this for some online activities where an email address is required, in order to avoid using their actual email address. For near-total online anonymity, programs like Anonymizer and Ghostsurf will hide one's IP address and erase browser information for a relatively low cost (\$80 and \$40 for one-year subscriptions, respectively, as of February 2011), although they apparently make the browsing process slower.

The market has provided these tools. How often consumers use these programs is unclear, but we would guess they are not used very often, and rationally so. Nonetheless, there are privacy solutions available to consumers who are truly concerned.

V. Conclusion

We have argued that the current system of personal information collection, targeted advertisements, and free online services and applications works very well in the United States. The critics of this system have done an insufficient job describing and quantifying the dangers that they fear. They have also mistakenly tried to address security problems as privacy problems, and generally seem reluctant to view the issue of online privacy in economic terms. Of course, there is more to life and policy than economics, but every policy decision involves costs and benefits, whether one recognizes them explicitly or not. We believe they should be made explicit, if possible.

The damage from identity fraud and the value to consumers of protecting their personal information are small relative the huge value provided by ad-supported online services. There is some evidence that Europe's stricter privacy regulation has reduced the value of targeted ads, which should, in turn, be expected to reduce funding for free online services. This does not prove that there are no privacy regulation changes that would create benefits greater than costs, but radical changes could upset the system that has created so much consumer surplus.

Surveys have repeatedly shown that consumers do not like targeted ads or the collection of personal information, and they suggest that consumers do not understand cookies or online privacy in general. They also suggest that most consumers who say they care about protecting their personal information fail to take basic steps to do so. We argue that their support for stricter privacy regulation is an example of rational irrationality. Politicians and regulators will certainly pay attention to the opinions of consumers—they ignore consumers and voters at their own peril—but that does not mean that the policy views of consumers are necessarily correct. ▼

WE ARGUE THAT THEIR SUPPORT FOR STRICTER PRIVACY REGULATION IS AN EXAMPLE OF RATIONAL IRRATIONALITY.

- 1 See A. Diana, *European Union to Update Privacy Laws*, INFORMATION WEEK (Nov. 4, 2010), available at http://www.informationweek.com/news/security/privacy/showArticle.jhtml?articleID=228200181&cid=RSSfeed_IWK_All.
- 2 See R. Singel, *FTC Backs 'Do Not Track' Browser Setting* (Dec. 1, 2010), WIRED, available at <http://www.wired.com/epicenter/2010/12/ftc-do-not-track/>.
- 3 See K. Walker, *The Costs of Privacy*, 25 HARV. J.L. & PUB. POL'Y 87, 89 (2001).
- 4 See B. Szoka & A. Thierer, *Targeted Online Advertising: What's the Harm & Where Are We Heading?*, 16 PROGRESS ON POINT 1, 3 (June 2009). Available at <http://pff.org/issues-pubs/pops/2009/pop16.2/targetonlinead.pdf>.
- 5 See T. Lenard & P. Rubin, *In Defense of Data: Information and the Costs of Privacy*, 2 POL'Y & INTERNET 149-183 (2010).

- 6 See R. Hahn & A. Layne-Farrar, *The Benefits and Costs of Online Privacy Legislation*, AEI-Brookings Joint Center Working Paper No. 01-14, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=292649.
- 7 See M. Rotenberg, *Statement for the Record of The Electronic Privacy Information Center (EPIC)*, Hearing on "Do Not Track Legislation: Is Now the Right Time?" available at http://epic.org/privacy/consumer/EPIC_Do_Not_Track_Statement_120910.pdf.
- 8 See M. Rotenberg, *Data Protection and Search Engines on the Internet: Google-DoubleClick and other case Studies*, available at http://epic.org/privacy/ftc/google/EPIC_LIBE_Submission.pdf.
- 9 See B. Givens, *Online Information Brokers: Where is the Balance*, presentation at Computers Freedom and Privacy Conference, San Jose, CA, San Jose State University (Jun. 15, 2010), available at <http://www.privacyrights.org/node/3438>.
- 10 R. Gellman, *Privacy, Consumers, and Costs: How the Lack of Privacy Costs Consumers and Why Business Studies of Privacy Costs are Biased and Incomplete* (2002), available at <http://epic.org/reports/dmfprivacy.html>.
- 11 See J. Chester & E. Mierzwinski, *Comments of The Center for Digital Democracy and U.S. PIRG before the Federal Trade Commission*, (Feb. 18, 2011), available at <http://www.democraticmedia.org/files/2011-02-17-FTC-Comments.pdf>.
- 12 See M. Rotenberg, *Testimony and Statement for the Record of Marc Rotenberg, Executive Director, EPIC*, Legislative Hearing on "H.R. 2221, the Data Accountability and Trust Act and H.R. 1319, the Informed P2P User Act," (May 5, 2009), available at http://epic.org/linkedfiles/rotenberg_house_ctcp2221_1319.pdf.
- 13 See M. Rotenberg and J. Verdi, *Comments of the Electronic Privacy Information Center*, (Apr. 28, 2008), available at http://epic.org/privacy/idtheft/042808_ftc.pdf.
- 14 See R. Anderson & R. Moore, *The Economics of Information Security*, 314 SCIENCE 610, 610-613 (2006).
- 15 T. Lenard & P. Rubin, *Much Ado About Notification*, 29 REGULATION 44, 44-50, (Spring 2006).
- 16 See S. Romanosky, R. Telang, & A. Acquisti, *Do Data Breach Disclosure Laws Reduce Identity Theft?* (Nov. 9 2010) SSRN Working Paper No. 1268926, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1268926.
- 17 See R. Anderson, R. Boehme, R. Clayton, and T. Moore, *Security Economics and European Policy*, MANAGING INFORMATION RISK AND ECONOMICS OF SECURITY (M. Johnson, ed.) (Springer 2009).
- 18 For discussion of liability for security failure, see Anderson et al. *Supra* note 14, and House of Lords, *Personal Internet Security*, 39 (Aug. 2007), available at <http://www.publications.parliament.uk/pa/ld200607/ldselect/ldstech/165/165i.pdf>; and for a theoretical analysis, see H. Ogut, N. Menon, & S. Raghunathan, *Cyber Insurance and IT Security Investment: Impact of Interdependent Risk*, 4th Workshop on the Economics of Information Security (June 2 to 3, 2005).
- 19 See *supra* note 12.
- 20 See S. Kessler, *Online Behavior Tracking and Privacy: 7 Worst Case Scenarios*, ABC NEWS (Nov. 6, 2010), available at <http://abcnews.go.com/Technology/online-behavior-tracking-privacy-worst-case-scenarios/story?id=12068346&page=4>.

- 21 See R. McRill, *Spokeo Privacy Violation Warnings Spread by Facebook Users*, Associated Content from Yahoo (Jan. 2, 2011), available at http://www.associatedcontent.com/article/6175590/spokeo_privacy_violation_warnings_spread.html.
- 22 See M Shiels, *Personal Data Could Become a Commodity*, BBC MOBILE (Oct. 19, 2010), available at <http://www.bbc.co.uk/news/technology-11571513>.
- 23 See Center for Democracy and Technology, *Online Behavioral Advertising: Industry's Current Self-Regulatory Framework is Necessary, But Still Insufficient On Its Own To Protect Consumers* (Dec 9, 2009), available at <http://www.cdt.org/policy/online-behavioral-advertising-industry%E2%80%99s-current-self-regulatory-framework-necessary-still-in>,
- 24 See J. Chester & E. Mierzwinski, *supra* note 11.
- 25 See Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009, Official Journal of the European Union, available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:En:PDF>.
- 26 R. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1-44 (1960).
- 27 See J. Bouckaert & H. Degryse, *Opt In Versus Opt Out: A Free-entry Analysis of Privacy Policies*, CESifo Working Paper Series No. 1831, CentER Discussion Paper No. 2006-96 (Sep. 2006), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=939511.
- 28 M. Staten & F. Cate, *The Impact of Opt-in Privacy Rules On Retail Credit Markets: A Case Study of MBNA* 52 DUKE L. J. 745 (2003) available at <http://www.law.duke.edu/shell/cite.pl?52+Duke+L.+J.+745>.
- 29 E. Johnson & D. Goldstein, *Do defaults save lives?*, 302 SCIENCE 1338 (2003).
- 30 R. THALER & C. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (2008).
- 31 Note that here we are treating security problems as privacy problems, for the sake of argument.
- 32 Javelin Strategy and Research, *The 2010 Identity Fraud Survey Report* (2010). There is a fee for the full report, but a summary is available at <https://www.javelinstrategy.com/news/831/92/Javelin-Study-Finds-Identity-Fraud-Reached-New-High-in-2009-but-Consumers-are-Fighting-Back/d.pressRoomDetail>.
- 33 Synovate, *Federal Trade Commission—Identity Theft Survey Report* (2003), available at <http://www.ftc.gov/os/2003/09/synovatereport.pdf> .
- 34 Synovate, *Federal Trade Commission—2006 Identity Theft Survey Report* (2006), available at <http://www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf>.
- 35 Javelin Strategy and Research, *The 2008 Identity Fraud Survey Report* (2010). The consumer version is available at https://www.javelinstrategy.com/uploads/files/803.R_2008_Identity_Fraud_Survey_Report_Consumer_Version.pdf.
- 36 The Census Bureau estimates business-to-consumer shipments were around \$288 billion in 2008. See U.S. Census Bureau, *E-Stats* (May 27 2010) available at <http://www.census.gov/econ/estats/2008/2008reportfinal.pdf>.
- 37 See Internet World Stats, *Internet Usage and Population in North America* (last accessed Feb. 26 2011) available at <http://www.internetworldstats.com/stats14.htm#north>.

- 38 See McKinsey and Company for IAB Europe, *Consumers Driving the Digital Uptake: The Economic Value of Online Advertising-based services for Consumers* (Sep. 2010) available at http://www.iab.net/media/file/White-Paper-Consumers-driving-the-digital-uptake_FINAL.PDF.
- 39 A 2009 IAB report found a wide range of estimates of the value of the ad-supported internet in the U.S., depending on how one calculated its value, and in some cases it is unclear whether the authors use "ad-supported internet" to refer to the entire internet, or just part of it. See J. Deighton, J. Quelch, and Hamilton Consultants, Inc. for IAB, *Economic Value of the Advertising-Supported Internet Ecosystem* (June 10, 2009), available at <http://www.iab.net/media/file/Economic-Value-Report.pdf>.
- 40 A. Goldfarb, & C. Tucker, *Privacy Regulation and Online Advertising*, (57) MGMT SCI. (2011).
- 41 See A. Westin, *How Online Users Feel About Behavioral Marketing and How Adoption of Privacy and Security Policies Could Affect Their Feelings* (Mar. 27 2008) available at <http://www.ftc.gov/bcp/workshops/privacyproundtables/Westin.pdf>.
- 42 See Gallup, *U.S. Internet Users Ready to Limit Online Tracking for Ads* (Dec. 21 2010), available at <http://www.gallup.com/poll/145337/Internet-Users-Ready-Limit-Online-Tracking-Ads.aspx>.
- 43 J. Turow, J. King, C. Hoofnagle, A. Bleakley, & M. Hennessy, *Americans Reject Tailored Advertising and Three Activities that Enable It*, SSRN Working paper (Sep. 29, 2009) available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214.
- 44 See S. Spoiekermann, J. Grossklags, & B. Berendt, *E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior*, Proceedings of the 3@footnote regular:rd ACM Conference On Electronic Commerce 38 (2001) available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=761107.
- 45 A. Aquisti & J. Grossklags, *Privacy and Rationality: A Survey*, PRIVACY AND TECHNOLOGIES OF IDENTITY: A CROSS-DISCIPLINARY CONVERSATION, 15-29 (K. Strandburg and D. Raicu, eds) (2006).
- 46 This is hard to believe. Anyone who has shopped online has surely used https, which relies on encryption. They probably did not understand what encryption meant.
- 47 See A. McDonald & L. Cranor, *An Empirical Study of How People Perceive Online Behavioral Advertising*, CyLab Technical Report 09-015, available at http://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab09015.pdf.
- 48 B. CAPLAN, THE MYTH OF THE RATIONAL VOTER (2007).
- 49 Caplan says that he chose the term *rational irrationality* to "emphasize both its kinship with and divergence from, rational ignorance."
- 50 P. RUBIN, DARWINIAN POLITICS (2002).
- 51 See C. NASS & C. YEN, THE MAN WHO LIED TO HIS LAPTOP: WHAT MACHINES TEACH US ABOUT HUMAN RELATIONSHIPS (2010), and B. REEVES & C. NASS, THE MEDIA EQUATION: HOW PEOPLE TREAT COMPUTERS, TELEVISION, AND NEW MEDIA LIKE REAL PEOPLE AND PLACES (1996, 2002).
- 52 The Media Equation, chapter 2, *supra* note 48.
- 53 On the other hand, consumer ignorance of how online privacy works could be simple rational ignorance—the benefits of being informed are nearly zero, while the costs of becoming informed can be high.

54 Note that we are not arguing that consumers cannot judge whether they are annoyed by targeted advertisements, or any sort of advertisements; clearly that is the sort of subjective consumer judgment that economists must respect.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Journalism, Competition, and the Digital Transition

Matthew Bye &
Oliver Bethell
Google

Journalism, Competition, and the Digital Transition

By Matthew Bye & Oliver Bethell**

This article considers the role of competition law and policy in shaping the news industry's digital transition. It begins by examining the shifting landscape for traditional media companies and describing Google's approach to news. The article then addresses arguments that exemptions from the antitrust laws are necessary to facilitate a digital transition by traditional news providers and concludes by considering some of the emerging business models that have been the subject of recent Department of Justice Business Review Letters.

| *Competition Counsels, Google

I. Introduction

The news industry is undergoing significant changes that present a number of challenges and opportunities. These challenges and opportunities have been driven in part by the proliferation of new technologies that are transforming the way people consume news and, as a consequence, view advertising associated with news content. Today, for example, both the news and the advertising that people view can be tailored specifically to what they are looking for. This simple fact has changed not only how news is delivered but also how users engage with it and how it is monetized.

The challenges facing traditional news providers' business model have triggered debate as to how antitrust and competition policy should apply to ensure the future of journalism. The Federal Trade Commission has been examining these issues in detail as part of a series of workshops it hosted in 2009 and 2010 under the title "How Will Journalism Survive the Internet Age?" The Commission subsequently released a Discussion Draft on "Potential Policy Recommendations to Support the Reinvention of Journalism" that outlined some possible regulatory and legislative solutions to support the news industry during their digital transformation.¹

This article considers the role of competition law and policy in shaping the news industry's digital transition. It begins by examining the shifting landscape for traditional media companies and describing Google's approach to news. The article then addresses arguments that exemptions from the antitrust laws are necessary to facilitate a digital transition by traditional news providers and concludes by considering some of the emerging business models that have been the subject of recent Department of Justice Business Review Letters.

II. Charting the Path Forward in a Digital World

A. CURRENT INDUSTRY CHALLENGES AND HOW BEST TO ADDRESS THEM

The current challenges faced by the news industry are business problems, not legal problems, and can only be addressed effectively with business solutions.² Indeed, these challenges, viewed in their historical context, simply reflect another inflection point for an industry that has faced periodic challenges to its business model as technology has evolved. For example, circulation by U.S. households has been on the decline since the early twentieth century; the number of newspapers distributed peaked between 1890 and 1920.³ Indeed, as Professor Jackaway has observed, with each communications innovation of the last 100 years, we have seen a repetition of the discussion that is taking place today over the future of journalism.⁴

The newspaper business is not immune from the truism that, in order to succeed, a business must respond to the demands of its consumers by delivering products and services that they want. For the news industry to adapt and thrive in the digital world, it must therefore first take into account how the internet has changed its ability to sell consumers a bundle that may be more than the consumer wants or needs. Google calls this the “atomic unit of consumption”—the basic form of content that consumers desire.⁵

In the news field, the structure of the internet has caused the unit of consumption for news to migrate from full newspapers to individual articles. This transition has had profound implications for traditional media companies because newspapers never made much money from news. They instead made money from special-interest sections on topics such as automotive, travel, and home & garden. These sections attract contextually targeted advertising, which is much more effective than non-targeted advertising. Someone reading the automotive section is likely to be more interested in cars than the average consumer, so advertisers will pay a premium to reach those consumers.

Traditionally, the advertising revenue from these special sections has been used to cross-subsidize core news production; in other words, the automotive and real estate sections pay for the Baghdad bureau. Today, internet users go directly to websites like Edmunds, Orbitz, Epicurious, and Amazon to look for products and services in specialized areas. Advertisers follow those eyeballs, which makes

THIS IS REVENUE THAT HAS BEEN LOST, HOWEVER, DUE TO A CHANGE IN THE CLASSIFIED ADVERTISING BUSINESS MODEL—A CHANGE THAT REFLECTS INCREASINGLY VIGOROUS COMPETITION.

the traditional cross-subsidization model that newspapers have used far less profitable. That cross-subsidization was possible only because the print format allowed newspapers to capture their audiences and keep them.

The FTC’s Discussion Draft laments that newspapers’ classified advertising revenue has fallen from \$19.6 billion in 2000 to \$6 billion in 2009.⁶ This is revenue that has been lost, however, due to a change in the classified advertising business model—a change that reflects increasingly vigorous competition. Indeed, the loss of classified advertising to Craigslist, eBay, and other online advertisers has nothing to do with copying or free-riding, and everything to do with the emergence of a new, more effective, and more efficient product in the marketplace. Government antitrust agencies would ordinarily regard such a situation as a cause for celebration because consumers are getting a better product at a lower price.

B. GOOGLE NEWS

As traditional media companies face increasing competition from online news sources and aggregators alike, some commentators have raised questions about Google News. The goal of Google News has always been to offer users the abili-

ty to search and access varied perspectives on a story in order to help them better understand current events.

To that end, Google indexes more than 50,000 sources in dozens of languages from around the world. News events of the day are identified and ranked by computer algorithms that reflect the publishing activity—the collective news judgment—of news organizations. Individual articles are then automatically selected and ranked based on factors such as freshness, location, relevance, and diversity of content, without regard to political viewpoint or ideology. Google News shows only a headline and sometimes a “snippet”—just enough for someone to decide if they’re interested in reading the story. Clicking on the link takes them directly to the publisher’s website. They do so at a rate of about one billion times a month from Google News alone.

Publishers have easy-to-use tools at their disposal to communicate instructions about whether they want search engines to index their sites, and Google’s policy is to respect those instructions. For example, using what is called the Robots Exclusion Protocol (“REP”) (which has been the de facto industry standard across the web for over 15 years), a site administrator who wishes to remove her website from Google’s index can easily do so using a “robots.txt” file. To remove sites or prevent search engines from crawling parts of a site, a webmaster may:

PUBLISHERS HAVE EASY-TO-USE TOOLS AT THEIR DISPOSAL TO COMMUNICATE INSTRUCTIONS ABOUT WHETHER THEY WANT SEARCH ENGINES TO INDEX THEIR SITES, AND GOOGLE’S POLICY IS TO RESPECT THOSE INSTRUCTIONS.

- **Use a “robots.txt” file to designate the content not to be indexed.** A robots.txt file enables site owners to restrict access to a website by search engine robots that crawl the web. A website owner can choose to block some pages or the entire site from Google’s web crawler by using a robots.txt file. If a website owner uses a robots.txt file to restrict access, Google will not crawl or index the content of pages blocked by the robots.txt file. However, Google may still index the website’s URL, if Googlebot finds those URLs on other pages on the web. As described below, Google will remove the website from its index if a noindex meta tag is present.⁷
- **Use a “noindex” meta tag.** When the Google crawler finds a website with a noindex meta tag on a page, Google will completely drop that page from its search results, even if other pages link to it. If the site is currently in Google’s index, Google will remove it the next time the crawler crawls the site. The meta tag allows the website owner to control access completely, on a page by-page basis.⁸

Through the use of the robots.txt file and the noindex meta tag, website owners are able to prevent their sites—or specific content on their sites—from being

indexed by Google's crawler. The control that a website owner has over the indexing of its content can be quite granular. Review sites can, for example, use robots.txt to control the number of reviews that Google can index while keeping individual destination pages in Google search results.

Notwithstanding the existence of these tools, some publishers have complained that excluding their content from Google News might have a negative effect on their natural search rankings. Such a complaint was recently investigated by the Italian Competition Authority, which ultimately found no infringement on behalf of Google. In addition, the Italian Competition Authority acknowledged that inclusion in Google News drives traffic to news publishers' sites and that this increased visibility has the potential to increase the revenues news publishers can obtain from online advertising. The authority also confirmed that the presence of publishers' content in Google News had no impact on the ranking of news publishers' sites in Google's natural search results.

The French Competition Authority similarly looked at news publisher complaints and concluded that Google had already taken steps to assuage their concerns. In addition, the French Authority noted that pay-walls and new devices offered news publishers new ways of monetizing content and that Google was actively participating in discussions with publishers regarding new revenue models.⁹

III. Antitrust Exemptions for News Organizations Will Harm Consumer Welfare

Rather than embrace the varied innovative revenue stream options possible through strategic partnerships, some traditional news publishers have seized on the concept of seeking blanket antitrust exemptions for collusive pricing behavior among newspapers as a path out of their difficulties. The FTC's Discussion Draft offered two antitrust exemption proposals: the first would allow news organizations to agree jointly to erect pay walls protecting their online content, and the second would allow news organizations to agree jointly on a mechanism requiring "news aggregators and others" to pay for the use of online content.¹⁰ These proposals amount to the same thing: allowing news organizations to coordinate on payment schemes, rather than compete fairly and innovate apace. Adopting either of these antitrust exemptions would be a mistake, both as a matter of law and public policy.

Historically, antitrust exemptions have been disfavored by government enforcement agencies and courts alike. Referencing the Sports Broadcasting Act of 1961, which offered antitrust exemption to certain NFL activities, Judge Easterbrook criticized such acts as "special interest legislation, a single-industry exception to a law [namely, the Sherman Act] designed for the protection of the public ... recognition that special interest legislation enshrines results, rather than principles, is why courts read exceptions to the antitrust laws narrowly, with

beady eyes and green eyeshades.”¹¹ The Supreme Court in *Associated Press v. United States* made it clear that newspapers should be subject to the same legal standards as other businesses: “All are alike covered by the Sherman Act.”¹² In fact, the Supreme Court explicitly rejected the argument that “newspapers are entitled to a different and more favorable kind of trial procedure than all other persons covered by the Sherman Act.”¹³

Previous iterations of antitrust exemptions in the newspaper industry have been, at best, ineffective and, at worst, actively harmful to consumers, by increasing advertising and circulation prices while enriching corporations who were not the intended beneficiaries of the legislation.¹⁴ For example, the Newspaper Preservation Act (“NPA”), which was passed in 1970 allowed newspapers to form a joint operating agreement (“JOA”) that collectively set circulation advertising rates if, among other things, they maintained separate editorial boards.¹⁵ The NPA offered antitrust immunity to certain JOAs that had been formed before its passage, as well as allowing JOAs for newspapers that were in probable danger of financial failure.¹⁶ The NPA ultimately favored large news organizations, putting smaller, emerging media companies at a distinct competitive disadvantage. Furthermore, the creation of these shared monopolies simply increased entry barriers, creating a further diminution of competition. The primary result of creating immunity from liability for jointly setting prices has been and would simply be high prices for consumers.

PREVIOUS ITERATIONS OF ANTITRUST EXEMPTIONS IN THE NEWSPAPER INDUSTRY HAVE BEEN, AT BEST, INEFFECTIVE AND, AT WORST, ACTIVELY HARMFUL TO CONSUMERS.

Perhaps the clearest repudiation of antitrust exemption policies was delivered in the Commission-authorized statement of Alden F. Abbott, the FTC’s then associate director for policy and coordination in connection with the Antitrust Modernization Commission’s consideration of statutory exemptions and immunities:

“Basic economic theory teaches that an unregulated competitive market generally leads to the economically efficient level of output. In contrast, a restraint that effectively raises price above the competitive level (or, equivalently, reduces output below the competitive level) generally will result in consumers purchasing less of the product or service, and firms producing less, at the higher price, than would be the case under competitive conditions. Consequently, such a restraint results in a decrease in economic welfare. Further, it is well accepted that competition itself is an engine that drives economic efficiency. Therefore, logic suggests that antitrust exemptions may well handicap the economic progress of industries they are intended to protect. Individual firms may enjoy the benefits of antitrust exemptions, but

consumers and the economy bear the harm—and the sheltered sector is rendered less efficient overall.”¹⁷

Abbott continues, explaining that even industries that currently enjoy antitrust exemptions do not need them to grant amnesty to “efficient, socially useful forms of conduct,” because such conduct should pass the antitrust test of reasonableness.¹⁸

News organizations, rather than seeking immunity for anticompetitive behavior, should instead work within the antitrust framework to establish payment schemes that allow them to benefit from their online content without engaging in price-fixing. As Christine A. Varney, the current Assistant Attorney General for the DOJ’s Antitrust Division recently noted: “Any new exemption from the antitrust laws seems particularly inappropriate at this point—industry dynamism should be given a full opportunity to play out in the marketplace before any antitrust exemption is even considered.”¹⁹

IV. Antitrust Does Not Impede Innovative Business Models

Traditional media companies need not fear antitrust laws if their proposed collaborations are, in fact, designed to yield pro-competitive user benefits. The antitrust laws do not seek to constrain newspapers from finding creative solutions to the challenges that competition from online sources brings. As the FTC and DOJ’s Guidelines on Competitor Collaborations make clear, courts and enforcement agencies take a nuanced approach towards a proposed joint venture’s overall competitive effect, asking whether competitive harm is threatened, whether any competitive benefits exist, and whether the benefits outweigh the harm.²⁰ Newspapers’ attempts to cut costs, improve service, lower prices, or offer new or better content through competitor collaborations are unlikely to yield intense antitrust scrutiny.

In the last year, the DOJ has issued two business review letters relating to newspaper collaborations that demonstrate the antitrust laws’ inherent flexibility. The first business review letter involved a proposal by MyWire Inc. to develop and operate an internet subscription news aggregation service called the Global News Service.²¹ MyWire’s plan for its Global News Service was to allow users to browse among both related free- and fee-based material from varying publishers’ sites in a single interface. The DOJ’s business review letter concluded that the vertical agreements reached by MyWire with content providers would benefit consumers by allowing them access to a broad array of related content without the need to conduct individualized searches. Publishers would also stand

to benefit from increased traffic and the revenue generated by the Global News Service subscriptions. In light of the non-exclusive nature of those vertical agreements and the independent pricing power retained by MyWire, the DOJ assured MyWire that it had no intention of challenging its proposal.

In the second instance, in April 2010, the DOJ issued a business review letter stating that it did not intend to challenge the Associated Press' proposal to develop and operate a voluntary, centralized news registry designed to facilitate the licensing and online distribution of news content created by the AP and other similarly situated news originators. The AP intended to give content owners the ability to control what content appeared in this registry, as well as dictate the terms by which such content could be licensed. In giving its tacit approval of the registry, the DOJ cited the granular control content owners could exercise over the news included in the registry as an important factor in their conclusion that the registry would not lessen competition. Additionally, the DOJ noted that the reduction of transaction costs for content owners who could determine licensing and terms through the registry was a significant pro-competitive benefit that might encourage the proliferation of licensed content to users. In the DOJ's judgment, the registry offered the "promise of a new efficient way for licensing and tracking news content over the Internet."²²

Christine Varney in her recent speech discussing the news industry cited these business letters as illustrative of "the latitude publishers have as they meet the demands of the twenty-first century media marketplace. Collaborations that do not restrain competition unnecessarily pass muster under the antitrust laws, particularly if those collaborations promise efficiencies or other benefits."²³

V. Conclusion

The Fourth Estate is too crucial a part of a functioning democracy, and the internet too powerful a medium, for journalism to die in transition to a web-first approach. Nonetheless, calls for further regulation or legislation that are designed to cushion this transition should be viewed with extraordinary caution and skepticism. Antitrust exemptions are rarely good for consumer welfare and, in this case in particular, are simply not necessary given the innovative business models that are emerging to facilitate digital transition in the news industry. ▼

-
- 1 The Discussion Draft offered proposals in four significant subject areas: namely, copyright; antitrust; tax and corporate innovations; and government partnership. This article discusses the proposals that relate to antitrust only. Fed. Trade Comm'n, FEDERAL TRADE COMMISSION STAFF DISCUSSION DRAFT: POTENTIAL POLICY RECOMMENDATIONS TO SUPPORT THE REINVENTION OF JOURNALISM (2010), <http://www.ftc.gov/opp/workshops/news/jun15/docs/new-staff-discussion.pdf> (hereinafter DISCUSSION DRAFT).
 - 2 Google's Comments on the Federal Trade Commission's News Media Workshop and Staff Discussion Draft on Potential "Policy Recommendations to Support the Reinvention of Journalism" are available at <http://googlepublicpolicy.blogspot.com/2010/07/business-problems-need-business.html>.

- 3 JOSEPH D. STRAUBHAAR & ROBERT LAROSE, *MEDIA NOW: UNDERSTANDING MEDIA, CULTURE, AND TECHNOLOGY*, at 94 (2005).
- 4 See GWENYTH L. JACKAWAY, *MEDIA AT WAR: RADIO'S CHALLENGE TO THE NEWSPAPERS, 1924-1939*, at 1-10 (1995).
- 5 See *The Future of Journalism: Hearing Before the Subcomm. on Commc'ns, Tech., & the Internet of the S. Comm. on Commerce, Sci., & Transp.*, 111th Cong. (2009) (statement of Marissa Mayer, Vice President, Search Products and User Experience, Google, Inc.).
- 6 DISCUSSION DRAFT, *supra* note 1 at 2.
- 7 See Google Webmaster Central, Block or Remove Pages Using a Robots.txt File, <http://www.google.com/support/webmasters/bin/answer.py?answer=156449>.
- 8 See Google Webmaster Central, Using Meta Tags to Block Access to Your Site, <http://www.google.com/support/webmasters/bin/answer.py?answer=93710>.
- 9 Avis n° 10-A-29 du 14 décembre 2010 sur le fonctionnement concurrentiel de la publicité en ligne, ¶¶ 383-385.
- 10 DISCUSSION DRAFT, *supra* note 1 at 13.
- 11 *Chi. Prof'l Sports Ltd. v. Nat'l Basketball Ass'n*, 961 F.2d 667, 671-72 (7th Cir. 1992).
- 12 *Associated Press v. United States*, 326 U.S. 1, 7 (1945).
- 13 *Id.* at 6 and 20.
- 14 DISCUSSION DRAFT, *supra* note 5, at 14 (citing Leonard Downie, Jr. & Michael Schudson, *The Reconstruction of American Journalism*, COLUM. JOURNALISM REV., Oct. 19, 2009, at 28-29, available at http://www.cjr.org/reconstruction/the_reconstruction_of_american.php?page=all; Robert G. Picard, Tremors, Structural Damage and Some Casualties, but No Cataclysm: The News about News Provision, Background Paper to the Presentation by the Author at the Federal Trade Commission Workshop: From Town Crier to Bloggers: How Will Journalism Survive the Internet Age? (Nov. 7, 2009), at 11, <http://www.ftc.gov/os/comments/newsmediaworkshop/544505-00029.pdf>).
- 15 Newspaper Preservation Act, Pub. L. No. 91-353, 84 Stat. 466 (1970).
- 16 15 U.S.C. §§ 1802(5) and 1803(b).
- 17 Alden F. Abbott, Assoc. Dir. for Policy & Coordination, Bureau of Competition, Fed. Trade Comm'n, Prepared Statement Before the Antitrust Modernization Commission on Statutory Immunities and Exemptions (Dec. 1, 2005), at 3-4, <http://www.ftc.gov/os/2005/12/051202statutory.pdf> (citations omitted).
- 18 *Id.* at 6-7.
- 19 Christine A. Varney, "Dynamic Competition in the Newspaper Industry," Remarks Prepared for the Newspaper Association of America, at 7 (March 21, 2011), available at <http://www.justice.gov/atr/public/speeches/268742.htm>.

- 20 See generally, FTC AND DOJ ANTITRUST GUIDELINES FOR COLLABORATIONS AMONG COMPETITORS (2000), available at <http://www.ftc.gov/os/2000/04/ftcdojguidelines.pdf>.
- 21 Letter from Christine A. Varney, Assistant Attorney Gen., U.S. Dep't of Justice, to Charles E. Biggio, Esq. (Feb. 24, 2010), available at <http://www.justice.gov/atr/public/busreview/255624.pdf>.
- 22 See Dynamic Competition in the Newspaper Industry, *supra* note 18. Similar analyses are being undertaken in other jurisdictions as well. The French Competition Authority, for example, recently looked at a so-called "digital newsstand" joint venture under consideration by news publishers in France ("Les projets de kiosques numériques"). They concluded that "such initiatives may well constitute pertinent responses to the concerns or questions raised above [and] may allow the press to establish a viable and long-lasting economic model for the Internet." See Avis n° 10-A-29, *supra* note 8 at ¶¶ 386-396.
- 23 *Id.*



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

The Antitrust Economics of Free

David S. Evans

Global Economics Group,
Univ. of Chicago, & UCL

The Antitrust Economics of Free

By David S. Evans

This article examines antitrust analysis when one of the possible subject products of an antitrust or merger is ordinarily offered at a zero price. It shows that businesses often offer a product for free because it increases the overall profits they can earn from selling the free product and a companion product to either the same customer or different customers. The companion product may be a complement, a premium version of the free product, or the product on the other side of a two-sided market. The article then shows how antitrust and merger analysis should proceed when the subject is either the free product or the companion product. A key point is that the existence of a free good signals that there is a companion good, that firms consider both products simultaneously in maximizing profit, and that commonly used methods of antitrust analysis, including market definition, probably need to be adjusted to properly analyze two inextricably linked products. When antitrust or merger analysis involves a free product, the analysis of consumer welfare and injury also needs to account for customers of both the free product and its companion product since any change in market conditions for customers of one product affects the customers of the other product. Much of the analysis of the article is also relevant to other common situations in which price is set less than marginal cost.

David S. Evans is Chairman, Global Economics Group; Lecturer, University of Chicago Law School; Executive Director, Jevons Institute for Competition Law and Economics, and Visiting Professor, University College London. He can be contacted at devans@globaleconomicsgroup.com. He would like to thank Howard Chang and Angela Zhang for helpful comments and discussion.

I. Introduction

Consumers can get many products and services at a price of zero. They do not have to pay money to use Adobe Flash; post a resume on Monster; watch the Super Bowl on Fox; pay with a Visa debit card; use Google’s search engine; post messages to their friends on Facebook; find businesses through the Yellow Pages; download many applications for their iPhones and iPads; or use the Linux operating system. It seems like “free” is a feature of modern times, but people have also historically paid zero prices for many products—for radio since the 1920s, for using general purpose payment cards since the introduction of those cards in 1950, and, going back millennia, for a man getting a bride from the village matchmaker.¹

Zero prices result in conundrums and confusion in antitrust analysis. The SSNIP test becomes inoperable when the basic price is zero. There is no sound way to analyze a 5 percent increase in a price of zero—5 percent of zero is still zero. The analysis of market definition and power therefore becomes a challenge with commonly used analytical tools. Companies sometimes argue that their product or service should not be subject to antitrust scrutiny because it is free. In *Kinderstart v. Google*² a U.S. federal court granted Google’s motion for summary judgment in part because the court concluded that it is not possible to have a relevant antitrust market for something that is given away for free. Chinese search engine Baidu made the same argument in *Renren v. Baidu* and was rebuffed by the Chinese court.³ More companies would, in my experience, pursue this argument if their economic experts did not refuse to endorse the zero-price antitrust exemption.

There are several reasons to spend some effort sorting out what to do when the sticker price is 0.0. Despite the observation that free has a long pedigree, zero-price offers seem to have exploded with the growth of the web-based economy. The companies offering these great deals are sometimes large global companies that are already in the sights of the antitrust authorities. A number of high-profile antitrust cases have involved free products, including browsers and media players in the various Microsoft cases,⁴ search engines in the various investigations and antitrust cases involving Google,⁵ and free open-source software in Oracle’s acquisition of Sun.⁶

It will prove increasingly challenging to get antitrust analysis right as more and more antitrust cases and mergers involve companies that offer products as zero prices. Based on my experience, there is a tendency on the part of companies, authorities, and courts to do more hand waving than serious analysis when they encounter products and services offered for free. While

BASED ON MY EXPERIENCE,
THERE IS A TENDENCY ON THE
PART OF COMPANIES,
AUTHORITIES, AND COURTS TO
DO MORE HAND WAVING THAN
SERIOUS ANALYSIS WHEN THEY
ENCOUNTER PRODUCTS AND
SERVICES OFFERED FOR FREE.

one solution to the conundrum brought by zero prices is to figure out some way to ignore them, investment in getting the analysis right is unquestionably worthwhile given the vast amount of consumer surplus that likely results from products and services offered for free.

This article examines the challenge to conventional antitrust analysis when one of the possible subjects of an antitrust or merger is ordinarily offered at a zero price.⁷ Proper analysis must begin by understanding why the provider has decided to charge a price of zero. Section 2 summarizes the main economic reasons. Then, in Section 3, the article explores how a good or service offered at a zero price should factor into antitrust and merger inquiries. Modern antitrust and merger analysis relies heavily on market definition and, in particular, the hypothetical monopoly test. Section 4 examines the implications for market definition and the monopoly test when a product of interest carries a zero price. Consumers, all else equal, would seem to get a great deal of consumer surplus from free goods and services. Just consider the value to global consumers of getting free search results. Section 5 considers the analysis of consumer welfare and consumer harm when one of the goods or services implicated in an antitrust or merger matter is priced at zero. Section 6 concludes and makes the observation that the analysis in the preceding sections is also relevant to other common situations in which price is ordinarily set at less than marginal cost.

II. Economic Reasons for Free

While we will see some exceptions below, most companies charge a price of zero because doing so allows them to make more money than charging a positive price. Charging nothing for a product or service enables them to make money, somehow, somewhere else.

A. COMPLEMENTARY PRODUCTS

The recognition that a zero price could be profit maximizing was made early in the 20th century in the analysis of pricing by a monopolist of complementary products.⁸ Two products are complements if a decrease in the price of one product increases the price of the other product. Consider a monopoly that produces two complementary products. As it searches for the profit-maximizing price the monopolist realizes that, as it raises the price of one product, it reduces the sales and possibly the profits coming from the other product. If widgets are highly complementary to gadgets, and if the elasticity of demand for widgets is very high, then increasing the price of widgets results not only in a great increase in the sales of widgets as well as a great loss of sales of gadgets. It could be that the profit-maximizing price involves giving widgets away and making the money from the gadgets.⁹ The result does, of course, depend a bit on a Goldilocks result—the degrees of complementarity and the elasticities of demand have to be just right for the optimal price to be zero.¹⁰

This result is often described not in terms of widgets and gadgets but with razors and blades. That has led to the business advice “give away the razors and sell the blades.” That example has some problems, as Professor Picker has argued.¹¹ If a razor manufacturer gives away the razor and makes the losses up on the blades, a competitor could sell the blades at a lower price since it does not need to absorb the losses on the razors. A razor manufacturer can make money from its free razor policy only if it can use patents, product design, or other devices to prevent consumers from buying from a competitor. This is a general, although hardly insolvable, problem for durable goods makers who might consider giving away the durable such as a copying machine to make money in the aftermarket for toner.

For the free complementary good strategy to work in practice, the seller must have some market power over the customer during her purchasing decisions for the not free product. Consider snacks at a bar. The bar could charge the customer for peanuts and pretzels. But most bars provide the snacks for free. The more snacks people consume, the more drinks they will buy. To make this strategy work, the bar should eject customers that bring in their own cheaper drinks to get the free snacks. Other situations in which people are provided something for free have a similar profile. To continue the food example, restaurants provide seating, water, utensils, bathrooms, and other services at no charge. Hotels provide basic television for free and some even provide free internet access. American airlines used to allow people to check as many bags as they wanted for free but that policy was abandoned along with the free peanuts.

Customers do not have to be literally captured in the short run for the free strategy to work with complementary goods. Over the last decade American banks have given customers “free checking accounts” in the expectation that the banks would earn fees from complementary services offered by the bank, such as direct deposit and savings.¹² The banks bet that enough customers will make enough use of the complementary services to offset the costs of providing the free services.

Free, as mentioned above, is a special case. Often, sellers of complementary products will price one of the products low, without going all of the way to zero. American movie theatres earn much of their profits not from the admission fee for seeing the movie, but from the sale of complementary beverages and snacks.¹³ Supermarkets reportedly sell some products at “low prices” that are complementary to other products. Consumers buy the cheap milk and then put other more expensive items in the basket. Casual observation suggests that setting the price of a complementary good exactly at zero is relatively rare. As of today, even the famous free razor has an implied price of several dollars.¹⁴

OFTEN, SELLERS OF
COMPLEMENTARY PRODUCTS WILL
PRICE ONE OF THE PRODUCTS
LOW, WITHOUT GOING
ALL OF THE WAY TO ZERO.

B. MULTI-SIDED PLATFORMS

A number of businesses are based on multi-sided platforms that serve two or more distinct groups of consumers, each of whom can provide a source of revenue.¹⁵ At least one of the consumer groups values being on the same platform as the other group of consumers. The profit-maximizing prices for each group depends on its level of demand, the interdependencies between itself and the other group, and possibly the marginal costs of producing the products.¹⁶ In some ways this is similar to the traditional analysis of complementary products. But here the complementary product for members of one group of consumers is the members of the other group of consumers. If the elasticities of demand and cross-dependencies between the demands of each group line up properly, it is possible that the profit-maximizing price for one of the products is zero.

While again, this is a Goldilocks condition, it turns out to be empirically quite important. A price of 0.0 is common across diverse industries, examples include:

- The general purpose charge card, introduced in 1950 in the United States. People value charged cards to the extent that merchants take them for payment, and merchants value accepting charge cards to the extent that they get incremental sales from accepting this form of payment. The card companies charge consumers a zero price for transactions and an annual fee that is largely, if not completely, offset by the float that consumers get.¹⁷ They charge merchants a percent of the transaction amount.
- Shopping malls have two groups of customers: the retailers who locate there and the consumers who shop there. Most malls do not charge consumers; shopping at a mall is usually free. The mall owners make their money from retailers.
- Microsoft Windows provides valuable services to both users who use it as their operating system and developers who write, and sell, applications for it. The developers get most of the benefits for free while the user pays (indirectly, in this case, to original equipment manufacturers who install Windows on machines that are sold to users).
- Online job boards such as Career Builder do not charge users anything to view job posts, but make their money from employers who are seeking to find workers.
- Advertising-supported media provides several examples. Google provides search engine services for free and makes its money from advertising. Facebook provides social networking services for free and makes its money from advertising and other complementary products such as games. Countless free newspapers, websites, radio stations, and free television stations provide content for a zero price and make their money from advertisers. OpenTable provides a restaurant reservation service to consumers for free; it charges participating restaurants,

which make their money from the patrons who have made their reservations through OpenTable.

There is nothing about the economics of multi-sided platforms that requires that customers on one of the sides are charged a zero price, or even a price below marginal cost.¹⁸ In fact, many multi-sided platforms earn significant revenues from both sides. Unlike Microsoft Windows on the PC, Apple, on the iPhone OS, not only charges applications developers 30 percent of their revenues, but also users (indirectly) for getting an iPhone or iPad. While many newspapers and magazines only charge subscription fees that roughly cover printing and distribution costs, others, such as *The Economist* and *People Magazine*, earn significant portions of their profits from both subscribers and advertisers.

The price structures for multi-sided platforms are not immutable. Magazines were mainly subscriber-supported in the 19th century United States. Many online newspapers, such as the *Wall Street Journal*, charge readers, and many more are starting to erect “pay walls,” eliminating the free-for-reader model they have relied on for many years. Nevertheless, as it happens, 0.0 is a common price for one side of many multi-sided platforms.

One important distinction between the multi-sided platform case and the complementary product case discussed above is that the beneficiaries of the subsidy are usually different. Bar flies get the free nuts but pay for the drinks. People who make restaurant reservations with OpenTable pay nothing to OpenTable. The fact that these multi-sided platforms involve different groups of customers has important implications for the analysis of consumer welfare, as we will see below.

THE FACT THAT THESE
MULTI-SIDED PLATFORMS
INVOLVE DIFFERENT GROUPS
OF CUSTOMERS HAS
IMPORTANT IMPLICATIONS
FOR THE ANALYSIS OF
CONSUMER WELFARE

C. PREMIUM UPGRADE STRATEGIES

A common business strategy in the internet economy is to offer a basic product for free, but then charge for premium versions of the product. In some cases this may simply reflect two-sided market pricing strategy. The company charges a zero price for a basic version to develop an installed base of users that are valuable to growing the other side of the platform. But it charges a positive price for enhanced versions of the platform to earn revenue from some of these users. Adobe has adopted this strategy for its Adobe reader. Consumers can get the basic Adobe reader for free; that increases the demand for people to buy software that writes Adobe files. But then Adobe charges people for enhanced versions of its reader software—for example, for versions that enable readers to highlight or comment on certain passages. This strategy has also become popular for online newspapers. The *Wall Street Journal* and *The Financial Times* provide limited free

access to content, but charge subscription prices for access to the full publication. It is a two-sided strategy because these free users are attractive to advertisers.

In other cases, the business strategy is to use the basic version to get people to try and learn about a product. Some fraction of these people will upgrade. This can be profitable if the marginal cost of offering the basic version is low, as is often the case for software or online media. The revenue from the upgrades to the premium version more than covers the fixed costs of creating the product. SugarCRM, for example, is a customer-relationship management software package that is provided under the open source model. Sugar CRM makes the “community edition” model available for free, but charges \$360 for the “professional edition.”

D. FREE SOFTWARE

Software has had a long history of being free. From the 1950s through the 1970s many software programs were distributed for free, and the notion of charging for software was controversial. Congress extended copyright privileges to software programs in 1974 and, as a result of court interpretations of that legislation, it became relatively easy for application developers to copy their works. While free software never literally went away, it started to make a significant comeback in the 1990s as a result of the open source movement. This movement involved developing an institutional structure that granted licensing arrangements to software developers, who were, in turn, required to distribute their program enhancements for free.

Open source has resulted in the development of many freely available software languages and software programs. The most famous of these is Linux, but almost every software category has open source competitors and, in some cases, these free programs have significant market shares. Over time, paid business models have sometimes developed around these free software packages, including ones based on selling add-on services (RedHat Linux), selling premium versions (SugarCRM mentioned above), obtaining ancillary revenues (the Firefox browser receives money from Google for using Google’s search engine which benefits Google, which then gets advertising revenue), or selling complementary products (IBM).¹⁹

Software developers wrote and gave their programs away for free before copyright protection because, once they had developed the program for their own purposes, it was costless to distribute it and, further, creating a popular software program could enhance a programmer’s reputation. They have continued to do so despite having copyright protection available. Many applications for the iPhone and Droid operating systems are available for free. A June 2010 survey found that 23 percent of iPhone’s applications were free as were 57 percent of Droid applications.²⁰

III. Antitrust and Mergers Involving Free Goods and Services

The previous section identified situations in which a profit-maximizing firm would charge a price of zero for a good or service, and documented anecdotally that this practice was hardly unusual. The question arises: Does the fact that the supplier doesn't charge for a product imply anything about whether the antitrust laws should apply to that product?

There are several possible reasons for concluding that the antitrust laws are not relevant to things that are given away. If a product, by its nature, is free, then there is no concern that business practices will result in consumers paying a higher price for the product. Without the prospect of consumer harm, there is no reason to care about that product.

One could also question whether the notion of a market is even meaningful for a free good. The product is not really sold since consumers can get it for free and, in some cases, it is just there for the taking. Since a relevant antitrust market is usually a prerequisite for an antitrust claim, there would be no basis for pursuing such a claim under this theory.²¹

Another possible argument is that businesses providing a free product are almost certainly making money from some other product. Antitrust analysis can, therefore, focus on the relevant market for the paid companion product and the impact on consumers of that product. In a merger, for example, we would be concerned about the impact of the consolidation on the increase in price for the paid twin.

A common problem with all of these justifications for a “free exemption” is that they focus on price. Price is only one dimension of competition. Although it is often convenient for economists to concentrate on price in economic models of business behavior, it is generally understood that price in these models subsumes all non-price measures of competition, including quality differences. However, while a merger or monopolistic practice may not affect whether a product is given away for free, it could very well affect such non-price dimensions as product attributes, service, and innovation. In fact, it is possible that a merger or monopolistic practice could have no material effect on the price of the twin paid product but still harm consumers substantially as a result of reductions in product quality or investments in product improvements and innovation. For example, a merger of web-based advertising supported properties could change incentives regarding how much privacy protection to give consumers.

HOWEVER, WHILE A MERGER OR MONOPOLISTIC PRACTICE MAY NOT AFFECT WHETHER A PRODUCT IS GIVEN AWAY FOR FREE, IT COULD VERY WELL AFFECT SUCH NON-PRICE DIMENSIONS AS PRODUCT ATTRIBUTES, SERVICE, AND INNOVATION.

The argument that free goods are not sold also does not make economic sense. Businesses still have to make decisions on how much to supply at a price of zero, and consumers still need to decide how much to demand given that they generally need to expend resources to obtain and consume these free products. In terms of competitive demand and supply, or the standard framework for a profit-maximizing firm setting price in the face of a downward sloping demand schedule, a “free price” simply means that the competitive market or the profit-maximizing firm sets a price of zero. Zero is just another number.

Two products that have been the subject of antitrust inquiries in many jurisdictions illustrate the debate over the relevance of a zero price: search engines and payment cards.

TWO PRODUCTS THAT HAVE BEEN
THE SUBJECT OF ANTITRUST
INQUIRIES IN MANY JURISDICTIONS
ILLUSTRATE THE DEBATE
OVER THE RELEVANCE OF A ZERO
PRICE: SEARCH ENGINES
AND PAYMENT CARDS.

Web search engines enable people to search vast quantities of data for free. Their twin paid product is usually advertising. Companies sell space on search results pages to advertisers usually based on an auction for the keywords that people use to find those results. In most countries, there is a dominant search engine

that has more than 60 percent of the shares of search and search-based advertising, and often more than 90 percent.²² Courts in the United States and China have addressed the relevance of free search when considering antitrust claims regarding search engines.²³

In a case brought in U.S. Federal District Court, *Kinderstart*, a website that focuses on providing content related to young children, claimed, among other things, that Google had lowered its rank—and thus reduced the likelihood it would appear on search engine results pages—in violation of Section 2 of the Sherman Act. In March 2007, the court dismissed the complaint for a number of reasons, including the fact that *Kinderstart* had failed to establish its claim that search is a relevant antitrust market. Key to the court’s conclusion was that search was freely provided.

“*KinderStart* has failed to allege that the Search Market is a “grouping of sales.” It does not claim that Google sells its search services, or that any other search provider does so. Rather, it states conclusorily that “[a]ny search engine must be free to the user because of past user experience and expectations with search engines and due to the preexisting governmental and technological policy of Internet freedom and Internet neutrality.” SAC ¶ 54. *KinderStart* cites no authority indicating that antitrust law concerns itself with competition in the provision of free services. Providing search functionality may lead to revenue from other sources, but *KinderStart* has not

alleged that anyone pays Google to search. Thus, the Search Market is not a “market” for purposes of antitrust law.” [emphasis added]

(The judge noted that KinderStart might have argued for a combined search results and search advertising market. We will return to this subject below.)

A Chinese court, in December 2009, reached the opposite conclusion on the relevance of “free” in a case brought by Renren, a web-based provider of medical information, against Baidu, the leading search engine provider in China.²⁴ Renren claimed that Baidu reduced its rank in order to coerce Renren to spend more on advertising with Baidu. The court ruled in favor of Baidu on the grounds that Renren had not shown that Baidu had a dominant position in a relevant market. However, in the course of its analysis, it rejected Baidu’s claim that search could not be a relevant market because it provided search for free. According to Zhang,²⁵

“The court was unpersuaded by Kinderstart and reasoned that although the search engine service was free, the service was closely tied to other products and services for which Baidu does requires payment. Unlike free public internet service, search engine service generates actual or potential profits from advertising and marketing. Therefore, whether a service is free is an irrelevant factor in evaluating the relevant market.”

In many countries, associations of bank-owned networks connect merchants that accept payment cards with banks that issue payment cards. These networks, sometimes in consultation with their member banks, set an “interchange fee” that a bank receives from a merchant when one of its cardholders uses the card for a purchase. Some competition authorities have concluded that setting the interchange fee results from coordinated behavior among horizontal competitors and is, therefore, a violation of the antitrust laws.

The European Commission concluded that MasterCard and Visa infringed Article 101 EU Treaty as a result of setting the interchange fee.²⁶ However, the Commission recognized that having a centrally set interchange is economically desirable and that a lower fee would be exempt under Article 101(3).²⁷

In the United States, merchants have claimed in a private lawsuit that MasterCard and Visa violated Section 1 of the Sherman Act as a result of setting an interchange fee. However, they appear to argue that it would not be a viola-

tion if MasterCard and Visa adopted a rule saying that merchants would not have to pay any discount off of the amount of the payment to the issuing bank—in other words, if they had established the same “on-par” payment as with the presentment of checks. On-par reimbursement is mathematically equivalent to an interchange fee of zero. Thus the argument hinges on the claim that setting a price of zero does not involve price-fixing while a positive price paid to the issuer does. It is easy in this case to see that this reasoning is spurious. The privately optimal interchange fee could involve a positive payment to the merchant when

A PRICE OF ZERO PROVIDES
A RED FLAG THAT THE TEXTBOOK
MODEL OF COMPETITION
AND STANDARD ANTITRUST
ANALYSIS DO NOT APPLY TO
THE PRODUCT IN QUESTION.

a cardholder pays with her card and a charge to the cardholder. Raising the fee from a negative amount to zero would harm the merchant.

The fact that a product is free is not, however, completely irrelevant to the practice of antitrust. A price of zero provides a red flag that the textbook model of competition and standard antitrust analysis do not apply to the prod-

uct in question. Almost certainly the proper antitrust analysis will need to consider the free product together with its companion moneymaking product. If the antitrust inquiry centers on a free product, then the analysis should be expanded to the other products provided by the firm that, in effect, subsidize the provision of the free product. Business practices related to the free product could result in benefits or costs for consumers of the companion money-making product. If the antitrust inquiry centers on a money-making product that has a free counterpart, the analysis should be extended to the free product for the same reason.

A free price also implies that traditional tools of economic analysis need to be used with care. Antitrust analysis often relies on the basic finding that prices tend to equal the marginal costs of production in competitive markets, and that deviations from marginal cost prices indicate market power. When a firm sells a product that is usually free, it cannot be operating in the sort of markets described in elementary models. It probably loses money on this product (assuming, as is usually the case, that it costs something to produce the product) and, if so, it must be selling another twin product at a price in excess of marginal cost—because only by making a profit on some other good can it sustain the losses involved in offering a free product. Therefore, the firm could earn a competitive rate of return overall even if it is selling a product at considerably more than marginal cost.

IV. Defining Markets When Products and Services Are Free

The purpose of market definition, and the related analysis of market power, is to understand the competitive constraints that can limit the ability of a firm to engage in behavior that harms consumers.²⁸ The fact that a product is sold for

free usually indicates there is a companion product and that the economics of those products are inextricably intertwined. Profit-maximizing firms do not provide products for free unless it helps them make money somewhere else. Formally, when a firm sets a price at zero, it is the result of a firm selecting the prices for several interrelated goods and finding that the profit-maximizing prices involve setting price equal to zero for one or more of those goods so long as at least one good is sold for a positive price.

The interdependency of complementary products has been recognized in after-market cases. These involve situations in which a company markets a durable good such as a printer and sells consumable products such as printer ink to purchasers of the durable goods. The primary and after-market products are complements. An antitrust analysis would not reach a reliable conclusion if it defined a market for the consumable product and ignored competitive constraints arising from the primary product.

The U.S. courts have generally recognized this. In *Kodak*, Kodak's motion argued for summary judgment on the grounds that competition in the primary market precluded monopoly pricing in the aftermarket, but the Supreme Court rejected that position.²⁹ However, the Court recognized that monopoly pricing in the aftermarket could occur only under special conditions. Lower court decisions applying *Kodak* typically grant summary judgment to the defendant unless the plaintiff can show: 1) there are high switching costs after purchasing the primary product; 2) consumers lack information to conduct lifecycle cost estimates when purchasing the durable good; and 3) the manufacturer engages in post-sale opportunistic conduct to exploit the installed base of users.³⁰

Under this analysis, the courts treat the provision of the durables and consumables as separate for the purposes of determining the relevant antitrust markets, and concentrate on the consumables market since that is usually the focus of the antitrust complaint. They then consider the role of the primary market in constraining behavior in the aftermarket. This approach can result in a sensible outcome when competition for the durable sale constrains the lifecycle price and therefore the aftermarket price as well.

The aftermarket cases illustrate a general proposition in antitrust. In terms of reaching the right answer a sensible market power analysis can cure all defects in a market definition analysis. If the market is defined too narrowly, then constraints, such as those coming from the provision of complementary products, can demonstrate that the firm at issue lacks the ability to engage in harmful behavior. If a market is defined too broadly, then an analysis of constraints can find that a firm could engage in harmful behavior even though it seems like a relatively small participant in the market. In that case, under case law there would

THE FACT THAT A PRODUCT IS SOLD FOR FREE USUALLY INDICATES THERE IS A COMPANION PRODUCT AND THAT THE ECONOMICS OF THOSE PRODUCTS ARE INEXTRICABLY INTERTWINED.

probably need to be a rethinking of the market boundaries. This result is not surprising since the analysis of market definition and market power are both really about identifying the set of competitive constraints that determine whether or not a firm can engage in harmful behavior with respect to its customers.³¹

Problems arise when courts or competition authorities reach conclusions based on market definition without considering market power analysis, or by conducting a perfunctory market power analysis. Doing so can lead to errors. False negatives can arise when the market definition for the paid product results in the conclusion that there is no problem, but the harm arises in the free product. For example, a competition authority may decide not to block a merger based on the finding that price of the paid product would not rise by a small but significant amount (“SSNIP”); had it considered the overall impact of the merger on both the paid and free products it might have found small but significant harm to consumers. False positives can arise when market definition for the paid product results in a conclusion that there is market power for the paid product but the analysis ignores the fact that competition results in the dissipation of that market power when the paid and free products are considered together. Aftermarket

cases that ignore the impact of competition in the primary market are likely to lead to false positives.

PROBLEMS ARISE WHEN COURTS OR COMPETITION AUTHORITIES REACH CONCLUSIONS BASED ON MARKET DEFINITION WITHOUT CONSIDERING MARKET POWER ANALYSIS, OR BY CONDUCTING A PERFUNCTORY MARKET POWER ANALYSIS.

Several approaches should be considered when an antitrust or merger analysis involves a free product or when a paid product has a twin free product. The simplest case concerns the situation in which the free and paid products are substitutes; this occurs when there is a basic free

product and a premium paid product or when free open source products compete with paid products. From a theoretical standpoint, the usual analysis of market definition and market power when there are differentiated products applies in this case. But analysts need to deal with practical problems that arise from the fact that one of the products has a price of zero. Market share calculations become problematic. Basing shares on the value of sales would not make sense since it would ignore the constraint coming from the free products; basing shares on unit sales does not take into account quality differences for which price is a common proxy. There is no good mathematical solution for this problem and qualitative and judgmental analysis becomes necessary.

When there are complementary free and paid products there are two alternative analytical approaches. Although it is not common practice, market definition could consider complementary products as part of the set of competitive constraints. That would be consistent with my view that the market definition analysis should identify the firms, products, and institutions that are the sources of competitive constraints on the firms and products under consideration.³² Both the complementary free and paid products would be considered together as part

of a business ecosystem that is relevant for the firms and products under consideration. Alternatively, the analyst could consider the role of complementary free or paid products in the analysis of market power. As noted above, this approach would also minimize errors so long as the market power analysis is done seriously, and is not an afterthought to market definition.

Similar observations apply for market definitions for multi-sided platform businesses. The preferred approach usually involves recognizing that competition takes place between multi-sided platforms, and that the market consists of these firms as well as other firms operating on either side that impose competitive constraints. The *Kinderstart* court seemed to recognize this as a possible approach when it said that, “*Kinderstart* might have argued that the Search Market and the Search Ad Market combine to form one market for antitrust purposes.” However, no U.S. court, to my knowledge, has defined a market consisting of multi-sided platforms that provide services to distinct groups of customers.

The other approach involves defining relevant antitrust markets separately for the free and paid sides of the platform, but then taking the interdependencies into account in the analysis of market power. Again, so long as this analysis is not abbreviated, it could lead to the same result. Errors are minimized so long as the market definition and market power inquiries consider the full set of competitive constraints, including those coming from both sides of the platform.

A practical implication of a price of zero is that some of the standard tools of market definition and market power analysis break down as a pure mathematical matter. Consider applying the hypothetical monopoly test to determine the relevant market that includes the free product. One cannot conduct a hypothetical percent increase in price because 5 percent of nothing is nothing, and because the nature of the product may be such that the hypothetical monopolist would still find it profit-maximizing to price at zero. Similarly, price-cost margins cannot be used for critical loss analyses or for assessing market power (technically the price-cost margin would involve division by zero).

The reason why these tools break down in the case of a price equal to 0.0 brings us back to where we started in this section. A free price indicates that the pricing of the product, and the overall analysis of competition, cannot be based on traditional models of firm behavior. The analyst must recognize that there is a twin product and deal explicitly with the relationship between the two. There is extensive literature on how to consider pricing and business relationships in the case of multi-sided platforms. However, as I have argued elsewhere, while it is technically possible to extend the hypothetical monopoly test to two-sided platforms, the challenges of implementing the SSNIP test empirically in two-sided markets are likely to be overwhelming in practice.³³

When an antitrust or merger analysis involves a product that is made available for free—or where the paid product in question has a twin product whose price

is zero—there is no substitute for carefully considering the economic interrelationships between these products and the overall competition between providers of the paired products or one or the other product.

V. Consumer Welfare and Harm Involving Free Products and Services

A basic implication of the existence of a free product is that there is a twin product that may or may not be consumed by the same consumers of the free product. The economic analysis of these paired products demonstrates that, since firms are usually jointly maximizing profits over both products, anything that affects the demand or supply of one of these products necessarily affects the demand and supply of the other product. By the same token, anything that affects consumer surplus³⁴ for one product is likely to affect consumer surplus for the other product. To understand how a business practice, or prohibiting a business practice, affects consumer welfare one needs to consider both products, and their interdependencies, together.

Unfortunately, the mechanical application of market definition to antitrust matters can prevent courts and competition authorities from considering the welfare of all of the consumers that are directly affected by a business practice or its prohibition. Courts and competition authorities, having defined a market, typically focus the rest of their analysis on that market. If a court or competition authority defines a market over one of the related product, but not over the other, then it will likely consider the impact of the practice only on the consumers in that market. For traditional products, this practice makes sense in

UNFORTUNATELY, THE
MECHANICAL APPLICATION OF
MARKET DEFINITION TO ANTITRUST
MATTERS CAN PREVENT COURTS
AND COMPETITION AUTHORITIES
FROM CONSIDERING THE WELFARE
OF ALL OF THE CONSUMERS
THAT ARE DIRECTLY AFFECTED
BY A BUSINESS PRACTICE
OR ITS PROHIBITION.

terms of conserving judicial and authority resources; it would be time consuming and distracting to weigh all of the indirect effects, outside of the market, on other markets. For twin products, this approach makes no sense given that consumers of the product not considered in the market will directly feel the consequences of a business practice, or its prohibition, for the product for which a market has been defined.

The interchange fee cases illustrate the issue. A reduction in interchange fees necessarily increases the prices that cardholders pay, since banks will pass some portion of the lost revenue from merchants on to cardholders. It also necessarily reduces the prices that merchants pay, since acquiring banks will pass on some portion of the increased revenue to merchants in the form of lower prices. Consumers could obtain a benefit that would offset their costs if merchants passed on some portion of their sav-

ings in the form of lower prices. Evans, Litan, & Schmalensee show that, for the U.S. debit card business, a dramatic reduction in interchange fees is likely to harm consumers, at least in the short run.³⁵ Cases filed by plaintiff merchants have defined or proposed merchant-facing markets and largely ignored the impact of the behavior, or modifications to it, on consumers.³⁶ In cases in the EU addressing whether interchange fees constituted price fixing, the focus was on whether collectively set interchange fees raised prices to merchants, but consideration was given to benefits on the cardholder side facilitated by the existence of interchange fees.³⁷

The same point applies to analyzing the impact of a merger. If it involves businesses that produce related free- and paid products then the assessment of the merger should consider the impact of the merger on consumers of both products, even if those consumers are different. That could result in prohibiting mergers that do not impose significant harm on the paid product but do on the free product, or letting mergers proceed that impose significant harm on one product but provide offsetting benefits on the other product.

VI. Conclusion

Free goods and services are increasingly common as a result of the continuing development of web-based multi-sided platform businesses. There is no reason why these goods should receive any antitrust exemption through, for example, concluding (as the *Kinderstart* court did) that there is no relevant antitrust market for a free good. At the same time, the existence of a free good in an antitrust or merger inquiry—either as the subject of the inquiry or as a companion product to the subject of the inquiry—should signal to analysts that they need to understand the market forces that result in the provision of these interrelated products and the decision to price one of them for free.

Many of the issues discussed in this article for free goods also apply to products that are provided at prices below the marginal cost of production. These goods, like free ones, are economically rational for a firm to provide only if there is a companion product whose price is in excess (perhaps well so) of marginal cost. The two-sided market literature provides guidance on how to deal with these situations, but the existence of free- and low-priced goods can arise for other reasons as well. This reinforces the point that analysts need to understand the economics of these businesses and apply economic tools, and modes of analysis, that are relevant. ▼

1 Paying a price of zero is not necessarily the same thing as getting something for free. That is because money is sometimes only one of the values that buyers and sellers exchange. When a consumer goes to “free” websites such as www.yahoo.com the website often inserts a cookie into the consumer’s machine and retrieves information that helps the website secure higher prices for advertisements or provides data the website can sell. Consumers do not pay for listening to the radio but they incur a

cost, perhaps, of listening to the advertisements. People do not have to pay for credit card transactions but some fraction of them will end up revolving their balances or incurring late charges. Some goods and services do seem literally free though. There is no obvious value that people give up when they decide to use Adobe Flash, the Linux operating system, or the Yellow Pages. For the purposes of this paper we treat all these situations as part of the antitrust economics of free.

- 2 KinderStart.com, LLC v. Google, Inc. ("KinderStart"), 2007 WL 831806 (N. D. Cal.).
- 3 Tangshan Renren Information Services Co. ("Renren") v. Baidu, Opinion by Beijing No. 1 Intermediate People's Court, Civil Case No. Yizhongminchuzi 845/2009. The transcripts of the case are available (only in Chinese) at: http://www.chinacourt.org/zhibo/zhibo.php?zhibo_id=1865.
- 4 United States Of America v. Microsoft Corporation, No. 00-5212, Appeals From The United States District Court For The District Of Columbia, June 28, 2001, Civil Action No. 98-1232 (CKK); *Commission Decision of 24.03.2004 relating to a proceeding under Article 82 of the EC Treaty (Case COMP/ C-3/37.792 Microsoft)*, Commission of the European Communities (April 21, 2004).
- 5 European Commission, *Antitrust: Commission Probes Allegations of Antitrust Violations by Google*, Brussels (November 30, 2010); *Google Faces Texas Ag Inquiry, Settles Privacy Suit*, REUTERS (September 3rd, 2010); *Italy Launches Antitrust Probe Of Google News*, LAW 360 (August 27, 2009); *French Mapmaker Takes Google Maps To Court*, RFI, (July 29, 2009), available at http://www.rfi.fr/actuen/articles/115/article_4550.asp; U.S. v. Google Inc., Case No: 1:11-Cv-0,0688; Tradecomet.Com LLC v. Google Inc., Case No.09-Cv-01400; *Google, Inc v. Mytriggers.Com, Inc.*, Case No. 09-Cv-14836.
- 6 See, *Mergers: Commission clears Oracle's proposed acquisition of Sun Microsystems*, Case No. 5529, Commission of the European Communities, Brussels, (January 21, 2010).
- 7 The case in which the price is ordinarily zero—that is, where \$0.0 is the long-equilibrium price for a product—is different than cases in which firms charge zero prices temporarily for promotional or predatory purposes.
- 8 See ROY G.D. ALLEN, *MATHEMATICAL ANALYSIS FOR ECONOMISTS*, 508 (1938).
- 9 This is a special case of pricing below marginal cost. See *Id.* at 359–362. For a general discussion, see Stephen K. Layson, *Multimarket Monopoly, Marginal Revenue and Profit Margins*, Working Paper.
- 10 Some of the examples discussed in this section could also be explained based on the theory of two-part tariffs where there is a fixed access charge for the good and then a variable charge based on the use of consumables which provides a proxy for the intensity of demand. It is possible that the optimal access charge is zero for the durable and positive for the consumable. See JEAN TIROLE, *THE THEORY OF INDUSTRIAL ORGANIZATION* (1988).
- 11 Randal C. Picker, *The Razors-and-Blades Myth(s)*, Olin Working Paper No. 532, September 13, 2010. Available at SSRN: <http://ssrn.com/abstract=1676444>. From the standpoint of the theory the razor blade example is complicated by the fact that there are dynamics in which the razor is a durable good and the blades are consumable goods; the possibility of competition in razors and blades that could make the give-the-razor-away strategy unprofitable; the existence of alternative reasons why a high price for the razor and low price for blades could be profitable; and, at least for the early years of razors, the possibility that the free razors was an investment in getting consumers to experience a new product.
- 12 To qualify for a free account a consumer has to keep a minimum balance. The account is therefore not really free since the consumer is giving up the opportunity cost of the use of these funds.

- 13 Although concessions account for only about 20 percent of gross revenues, they represent some 40 percent of theaters' profits. That is because while ticket revenues must be shared with movie distributors, 100 percent of concessions go straight into an exhibitor's coffers. See Richard Gil & Wesley R. Hartmann, *The Role and Determinants of Concession Sales in Movie Theatres: Evidence from the Spanish Exhibition Industry*, 30(4) REV. INDUS. ORG., 325-347 (February 2007).
- 14 At cvs.com on April 10, 2011, the price of eight Gillette Mach 3 blades was \$21.99 for an average price of about \$2.75 while the price of a razor and two blades is \$9.29. The implied price of the razor is therefore \$3.79. For the economics of product bundling such as this see Michael A. Salinger & David S. Evans, *Why Do Firms Bundle and Tie? Evidence from Competitive Markets and Implications for Tying Law*, YALE J. ON REGULATION, (2004). Available at SSRN: <http://ssrn.com/abstract=550884>.
- 15 See E. Glen Weyl for one of the most recent analyses of this topic, *A Price Theory of Multi-Sided Platforms*, 100(4) AMER. ECON. REV., 1642-1672 (2010).
- 16 The precise relationship depends on the model. For the seminal paper in this area and one version of the optimal pricing relationships, see Jean-Charles Rochet & Jean Tirole, *Cooperation among Competitors: Some Economics of Payment Card Associations*, 33(4) RAND J. ECON., pp. 549-570 (Winter, 2002).
- 17 DAVID S. EVANS & RICHARD SCHMALENSEE, *PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING*, 2nd Ed. (January 1, 2005).
- 18 David S. Evans & Richard Schmalensee, *Markets with Two-Sided Platforms*, (1) ISSUES IN COMPETITION L. & POL'Y, Ch. 28 (2008), available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1094820
- 19 For a discussion of some of the reasons behind volunteering for open source software, see Josh Lerner & Jean Tirole, *Some Simple Economics of Open Source*, 50 (2) J. INDUS. ECON, pp. 197-234 (Jun., 2002).
- 20 See <http://techcrunch.com/2010/07/05/distimo-june-2010/>.
- 21 Of course neither of these arguments would apply to a predatory pricing case where free is a temporary situation. These are mainly arguments for products that we expect will always be free.
- 22 According to StatCounter, as of 2010, Google was the dominant search engine in most countries, with search shares of 80.9 percent in the United States, 92.5 percent in the United Kingdom, 95.6 percent in France, 97.1 percent in Germany, 92.7 percent in Canada, 95.1 percent in Australia, and 78.8 percent in Japan. Baidu was the leading search engine in China with a 60.9 percent share of search. These statistics are available at <http://gs.statcounter.com>.
- 23 Also see Renren, *supra* note 3.
- 24 For a discussion of this case, see *Id.* and R. Ian McEwin & Corinne Chew, *China—The Baidu Decision*, 6(2) COMPETITION POL'Y INT'L (Autumn 2010).
- 25 See Angela Huyue Zhang, *Using A Sledgehammer to Crack A Nut: Why China's Anti-Monopoly Law was Inappropriate for Renren v. Baidu*, 7(1) COMPETITION POL'Y INT'L (forthcoming, Spring 2011).
- 26 European Commission, *Antitrust: Commission Prohibits Mastercard's Intra-Eea Multilateral Interchange Fees*, (2007); and *Commission Plan To Clear Certain Visa Provisions, Challenge Others*, (2000).
- 27 European Commission, *Summary of Commission Decision of 19 December 2007 relating to a proceeding under Article 81 of the EC Treaty and Article 53 of the EEA Agreement*, (Case COMP/34.579

- MasterCard, Case COMP/36.518 – EuroCommerce, Case COMP/38.580 – Commercial Cards); and 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 No COMP/29.373 – Visa International – Multilateral Interchange Fee.
- 28 David S. Evans, *Lightening Up Market Definition*, RESEARCH HANDBOOK ON THE ECONOMIC OF ANTITRUST LAW, (Einer Elhauge, ed.), (forthcoming 2011). Available at SSRN: <http://ssrn.com/abstract=1599270>.
- 29 Eastman Kodak Co. V. Image Tech. Svcs., 504 U.S. 451 (1992) , No. 90-1029.
- 30 See David A.J. Goldfine & Kenneth M. Vorrasi, *The Fall Of The Kodak Aftermarket Doctrine: Dying A Slow Death In The Lower Courts*, (1) ANTITRUST L.J. (2004).
- 31 Howard H. Chang, David S. Evans, & Richard Schmalensee, *Market Power: Assessment of Market Power in Competition Matters*, and *Market Definition: Assessment of the Relevant in Competition Matters*, prepared for the Federal Competition Commission of Mexico, March 30, 2011.
- 32 Evans, *supra* note 28.
- 33 David S. Evans, *Two-Sided Market Definition*, MARKET DEFINITION IN ANTITRUST: THEORY AND CASE STUDIES, (Forthcoming). Available at SSRN: <http://ssrn.com/abstract=1396751>.
- 34 All the points here would be equally true if we focused on social surplus instead of consumer surplus.
- 35 For an analysis of the net impact on consumer welfare, see David S. Evans, Robert E. Litan, & Richard Schmalensee, *Economic Analysis of the Effects of the Federal Reserve Board's Proposed Debit Card Interchange Fee Regulations on Consumers and Small Businesses* (February 22, 2011), available at SSRN: <http://ssrn.com/abstract=1769887>.
- 36 In Re VISA Check/Mastermoney Antitrust Litigation, Master File No: CV-96-5238, May 26, 1999; and In Re Payment Card Interchange Fee Antitrust Litigation, Master File No. 1:05-md-1720-JG-JO, April 24, 2006.
- 37 European Commission, *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 No COMP/29.373 – Visa International – Multilateral Interchange Fee; and European Commission, *Summary of Commission Decision of 19 December 2007 relating to a proceeding under Article 81 of the EC Treaty and Article 53 of the EEA Agreement*, Case COMP/34.579 – MasterCard, Case COMP/36.518 – EuroCommerce, Case COMP/38.580 – Commercial Cards.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Are “Closed Systems” an Antitrust Problem?

Hanno F. Kaiser
Latham & Watkins LLP

Are “Closed Systems” an Antitrust Problem?

*By Hanno F. Kaiser**

Closed computer systems have come under attack as harmful to freedom, innovation, and competition. Open computer systems, in contrast, are said to promote such values. This article assesses the specific claim that closed systems, compared to open systems, are inherently anticompetitive. It concludes that competition policy arguments against closed systems are at best inconclusive and that closed systems should not be put in an antitrust suspect class.

*Partner, Latham & Watkins LLP, San Francisco. The views in this article are my own, so please do not impute them to my firm or my clients. This article is licensed under a Creative Commons Attribution 3.0 license.

I. Civic and Economic Criticisms of “Closed Systems”

In his influential book *The Future of the Internet and How to Stop It*¹ Jonathan Zittrain argues that we are headed for a future in which general purpose computers, which he calls “generative systems,” will be replaced by locked-down, tethered computing appliances:²

“The PC revolution was launched with PCs that invited innovation by others. So, too, with the Internet. Both were generative: they were designed to accept any contribution that followed a basic set of rules (either coded for a particular operating system, or respecting the protocols of the Internet. . . . But the future unfolding right now is very different from the past. The future is not one of generative PCs attached to a generative network. It is instead one of sterile appliances tethered to a network of control.”³

An important civic virtue of generative systems is that they invite and require participation.⁴ Much like open source software, generative systems do not draw sharp lines between consumers and producers, users and developers, because the tools of production are available to all. As a participant in a generative, open system, every user is a potential developer, much as every citizen in a democratic society is a potential lawmaker.

Zittrain is not alone in his critique of closed systems. In *The Master Switch*, Tim Wu chronicles the history of the information sector in the United States since the birth of AT&T. According to Wu, “[h]istory shows a typical progression of information technologies: . . . from a freely accessible channel to one strictly controlled by a single corporation or cartel—from open to closed system.”⁵ Open systems promote “a world in which most goods and services are free or practically free, thereby liberating the individual to pursue self-expression and self-actualization as an activity of primary importance.”⁶ Closed systems, in contrast, appeal to the consumer, not the creator. They are built to control the users, not to empower them, steering users towards “mass conformity.”⁷

Critics of closed systems generally view open systems as inherently superior in economic terms as well. For example, Zittrain argues that “[g]enerative systems facilitate change” specifically in the form of disruptive innovation.⁸ Disruptive innovation is commonly triggered by amateurs,⁹ who are not constrained by the business imperative of having to make a profit.¹⁰ Because

CRITICS OF CLOSED SYSTEMS
GENERALLY VIEW OPEN SYSTEMS
AS INHERENTLY SUPERIOR
IN ECONOMIC TERMS.

of their disruptive nature, generative PCs and the Internet “overwhelmed their respective proprietary, non-generative competitors: PCs crushed stand-alone word processors and the Internet displaced such proprietary online services as CompuServe and AOL.”¹¹ Timothy Berners-Lee agrees: “[C]losed, ‘walled gardens,’ no matter how pleasing can never compete in diversity, richness and innovation with the mad, throbbing Web market outside their gates.”¹² Jonathan Rosenberg similarly claims: “At Google we believe that open systems win. They lead to more innovation, value, freedom of choice for consumers, and a vibrant, profitable, and competitive ecosystem for business.”¹³

SOME CRITICS GO ONE STEP FURTHER, NOT MERELY CLAIMING THAT CLOSED SYSTEMS ARE LESS COMPETITIVE THAN OPEN SYSTEMS, IN WHICH CASE WE COULD REASONABLY EXPECT THE MARKET TO TAKE CARE OF THEM BUT, MORE SPECIFICALLY, ANTICOMPETITIVE.

Some critics go one step further, not merely claiming that closed systems are less competitive than open systems, in which case we could reasonably expect the market to take care of them but, more specifically, anticompetitive. As such, closed systems undermine the corrective dynamism of the market, which, in turn, justifies regulatory intervention. Wu, for example,

singles out Apple as the company he “fear[s] the most,”¹⁴ and that “despite the attention to Google’s monopoly, . . . is likely to run into antitrust problems first.”¹⁵ According to Wu, “unreasonably exclusionary” in the context of a monopolization offense “translates readily to a single word: ‘closed.’” In contrast, an open systems strategy “translates in antitrust language to ‘non-exclusionary.’”¹⁶ Wu approvingly cites Tom Conlon, who puts it more bluntly: “Once we replace the personal computer with a closed-platform device . . . , we replace freedom, choice and the free market with oppression, censorship and monopoly.”¹⁷

This article examines the specific claim about the supposed anticompetitive properties of closed systems.¹⁸

II. What’s Open, What’s Closed? Easy Labels Don’t Provide Ready Answers

What is a closed system? In antitrust law and economics, “systems” are often thought of as “collections of two or more components together with an interface that allows the components to work together.”¹⁹ Examples include applications and operating systems, nuts and bolts, video games, and gaming consoles. Often the components have little value in isolation but substantial value when combined with complementary system components (e.g., DVDs are useless without a DVD player and vice versa).

It is less clear what makes a system closed as opposed to open. First, there is a problem with the definitional scope of “open” and “closed.” A fully open system is an oxymoron because systems are, by definition, different from their environ-

ment and must therefore be closed in some respects.²⁰ If there is no locus of differentiation, i.e. if all parts of system A and system B are fully interchangeable, then neither A nor B is a system in any meaningful sense. Similarly, it is hard to come up with an example of a fully closed system, because even the most locked-down, tethered appliance must at least connect with the power grid. Open versus closed is therefore not a binary distinction but a matter of degree. All real-world systems are open in part and closed in others.

Second, computer systems or stacks consist of various layers—hardware, operating system, software, and content—each of which can be more or less open. Should a system like Microsoft Windows that is open at the content and software layers but closed at the operating system layer be labeled open or closed? Zittrain²¹ and Wu say “open,”²² the Free Software Foundation says “closed.”²³ What about platforms such as the Macintosh that are open at the software layer but closed at the operating system and the hardware layers? Zittrain says “open,”²⁴ Wu, Farrell, and Weiser say “closed.”²⁵ Everyone seems to agree that the Kindle, iPod, and TiVo are closed,²⁶ even though they all depend entirely on third party contributions—i.e., openness—at the content layer. Lastly, there is the vexing case of the iPhone/iPad. According to Zittrain, the iPhone was closed from June 2007 to February 2008.²⁷ After that, it turned into a “hybrid system.”²⁸ According to Wu, the iPhone remains “closed” to this day, despite the fact that there are over 100,000 iPhone developers who have created more than 300,000 applications, resulting in over 10 billion downloads.²⁹

As an analytical tool the labels “open” and “closed” are of limited utility, because they cannot adequately capture the complexity of selective openness at various layers of a system within their single binary distinction. Addressing the central antitrust issue requires that we move past the “ready labels” and focus on whether specific vertical restraints at all levels result in anti-competitive exclusion and foreclosure.

ACCORDING TO WU, THE IPHONE REMAINS “CLOSED” TO THIS DAY, DESPITE THE FACT THAT THERE ARE OVER 100,000 IPHONE DEVELOPERS WHO HAVE CREATED MORE THAN 300,000 APPLICATIONS, RESULTING IN OVER 10 BILLION DOWNLOADS.

III. The Treatment of Vertical Integration, Vertical Restraints, and Refusals to Interconnect

Vertical arrangements have a long, stormy, and well-documented history in antitrust law and economics. Until the late 1970s, courts and agencies were generally hostile towards vertical arrangements.³⁰ Modularity and open market structures, in which “[m]any firms compete in selling their individual components”³¹

at all levels of production, were the normative ideal. Against that baseline, closed organizations, in which “a single firm, or a small set of firms working hand in glove with one another, undertakes all those activities,”³² were suspect. The hostility towards vertical arrangements led to highly restrictive merger decisions such as *Brown Shoe*³³ or the *per se* illegality of intra-brand territorial distributor restraints.³⁴

Of course, the courts were correct in recognizing undeniable benefits from modularity, such as the ever more efficient production of components and resulting lower prices, superior component performance, and the gradual erosion of entry barriers.³⁵ However, focusing on the benefits of modularity to the exclusion

IF ANTITRUST CONCERNS REGARDING CLOSED SYSTEMS ARE REALLY CONCERNS ABOUT THE COMPETITIVE EFFECTS OF VERTICAL ARRANGEMENTS, AND VERTICAL ARRANGEMENTS ARE AFFORDED GREAT LENIENCY BY THE COURTS, WHAT IS THE BASIS FOR THE RECURRING CRITICISM OF CLOSED SYSTEMS AS POTENTIALLY ANTICOMPETITIVE?

of considering its costs left the courts with an incomplete picture. In its path-breaking 1977 *Sylvania* decision, the Supreme Court acknowledged that modular market structures are also vulnerable to systemic market failures stemming from transaction costs, lack of coordination, opportunism, free-riding, and double marginalization among others.³⁶ Once courts in the 1980s started taking both the benefits and the costs of modularity into account, it became clear that restrictions on access and other vertical restraints are usually a response to such market failures and only in exceptional circumstances driven by aspirations of anticompetitive exclusion.³⁷

Modern antitrust doctrine therefore recognizes that vertical restraints as such are not in an antitrust suspect class. To the contrary, most vertical restraints are pro-competitive, or else the marketplace would punish the firm imposing them. Only if the incumbent has significant market power and the restraint results in meaningful foreclosure can vertical restraints possibly have anticompetitive effects. As a consequence, courts today almost universally apply the rule of reason to all vertical restraints.³⁸

That leaves us with a puzzle. If antitrust concerns regarding closed systems are really concerns about the competitive effects of vertical arrangements, and vertical arrangements are afforded great leniency by the courts, what is the basis for the recurring criticism of closed systems as potentially anticompetitive? Before we can answer that question, we need to address one further complication. Computer platforms, after all, are not quite like nuts and bolts or restraints on bicycle distributors. There is something special about orchestrating the interactions of millions of users—and tens if not hundreds of thousands of hardware- and software developers, content providers, etc. What is it?

IV. Computer Platforms: System and Component Competition

On February 8, 2011, Stephen Elop, CEO of Nokia, sent an email to his employees, shortly before entering into a broad-based alliance with Microsoft to deploy the Windows Mobile operating system on Nokia handsets:

“The battle of devices has now become a war of ecosystems, where ecosystems include not only the hardware and software of the device, but developers, applications, ecommerce, advertising, search, social applications, location-based services, unified communications and many other things. Our competitors aren’t taking our market share with devices; they are taking our market share with an entire ecosystem. This means we’re going to have to decide how we either build, catalyse or join an ecosystem.”³⁹

Elop’s analysis reflects a shift in focus of the competitive interaction from devices to platforms (or ecosystems) and, further, a realization that platforms consist of broad coalitions of participants. Specifically, computer platforms generally consist of (1) the system sponsor, (2) various contributors, and (3) the users.⁴⁰ The system sponsor often contributes, among other things, the underlying platform technology (e.g., the operating system), developer tools, services to facilitate platform transactions (e.g., hosting and billing), platform governance, and related IP.⁴¹ Contributors include hardware and software developers that create platform-specific products, content providers (e.g., publishers, record labels), and services (e.g., Facebook, Twitter, Foursquare). In many instances, the platform sponsor also contributes to the platform by selling its own applications. The users, finally, include buyers of devices, applications, and services and everyone who uses indirectly monetized services on a given platform for free (e.g., Google or Bing search users).

There are two relevant realms of competitive interaction that are connected yet conceptually distinguishable: competition *among* ecosystems and competition *within* a given ecosystem, in other words, inter- and intra-system competition. System sponsors are the primary agents of inter-system competition (sponsor A versus sponsor B). Contributors are the primary agents of intra-system competition (e.g., developer A versus developer B). Problems arise if sys-

THERE ARE TWO RELEVANT REALMS OF COMPETITIVE INTERACTION THAT ARE CONNECTED YET CONCEPTUALLY DISTINGUISHABLE: COMPETITION AMONG ECOSYSTEMS AND COMPETITION WITHIN A GIVEN ECOSYSTEM.

tem sponsors wear multiple hats. In addition to being system sponsors, they may be application developers, hardware vendors, or service providers. This raises the familiar issue of non-integrated single-level competition versus vertically integrated multi-level competition. For example, complaints of platform sponsors giving their own applications an edge through “private” operating system application programming interfaces (“APIs”) or more powerful development tools fall into this category.⁴²

This concern, however, is limited to intra-system competition, because applications written for system A generally do not compete head-to-head with applications for system B, irrespective of whether the developer is vertically integrated or not.⁴³ That said, vertical arrangements within a given platform are not irrelevant in the inter-system context—far from it. Vertical integration within platform A can be among the most important competitive differentiators vis-à-vis platform B if it improves the overall value of the platform for all constituents, including the users. Examining the dual nature of intra-platform rules, promulgated by the platform sponsor, and their simultaneous impact on inter-system competition is key to our assessment of platform competition. It is to those effects that we now turn.

V. Intra-platform Rules with Inter-platform Effects

Every platform needs rules, if only to define the boundary between the system and its environment. Packets sent across the Internet must adhere to the TCP/IP protocols for delivery. In order to write an application for Linux, one must compile it to the Linux APIs. That much is uncontroversial. Some rules, however, go beyond open and well-documented technical standards and more directly limit participation in a given platform. For example, the online job board TheLadders only lists jobs that pay \$100,000 or more and charges job seekers a \$35 fee for applying to a job listing, sharply departing from the industry norm that the job seeker side of the platform is free.⁴⁴ TheLadders uses price as a means to limit user participation. Similarly, Microsoft requires would-be game developers for the Xbox or the Xbox Live Arcade to go through a multi-stage approval process before granting permission to distribute a game.⁴⁵ Why would a multi-sided platform

GIVEN THE PRESENCE OF
INDIRECT NETWORK EFFECTS,
ISN'T MORE ALWAYS BETTER?

limit participation or contributors and users? Given the presence of indirect network effects, isn't more always better?

At first blush, it appears so. Recall that multi-sided platforms generate value by facilitating transactions between at least two constituencies that stand to gain from interacting with one another. Internet search engines are a good example. Google's Jonathan Rosenberg summarizes the business model as follows:

“Users go where the information is so people bring more information to us. Advertisers go where the users are, so we get more advertisers. We get more users because we have more advertisers because we can buy distribution on sites that understand that our search engine monetizes better. So more users more information, more information more users, more advertisers more users, more users more advertisers, it’s a beautiful thing, lather, rinse, repeat, that’s what I do for a living.”⁴⁶

The European Commission adopted the “more is always better” reasoning in its 2010 *Microsoft/Yahoo!* decision and observed that “[i]n order to be successful, a search engine operator will try to attract as many participants on both sides of the platform as possible.”⁴⁷ Why then would any platform sponsor impose rules to *limit* participation—if not for some anticompetitive purpose?

The answer is that some platforms are much more sensitive than Internet search engines to negative externalities created by too many or by the wrong participants. The potential for market failure, in other words, is significant.⁴⁸ Consider the effects of crowding in a two-sided, low-tech platform setting.⁴⁹ An overcrowded nightclub is no fun for male or female patrons—the two constituencies that the venue seeks to connect. To minimize the resulting externalities (as Yogi Berra put it: “No one goes there anymore. It’s too crowded”), nightclubs limit admission and enforce that rule at the door. Now consider the impact of low quality contributions, the negative effects of which are not fully internalized by the contributor. Drunken or rowdy guests spoil the fun for both sides in a nightclub, which is why disorderly patrons are either denied access or are removed once they start being obnoxious. That too is a rule promulgated and enforced by the platform sponsor.

The same principles apply to computer platforms. As a rule, the developers go where the users go, and the users go where the developers go. The resulting positive feedback effect was the basis for the court’s finding of an “application barrier to entry” in the *U.S. v. Microsoft* case.⁵⁰ However, if there are too many applications in the same category, choosing the right one can become costly and time consuming for the users.⁵¹ The incremental search costs for users resulting from crowding on the developer side can be significant.⁵² In addition, if users cannot readily judge the quality of an application before buying it, then they may be unwilling to pay more than an average price for any application. As a result, low-quality contributions by some developers drive down the price for all applications, reduce the value of the platform for both users and competing developers, and diminish the developers’ incentives to invest in high-quality applications. A downward spiral ensues.⁵³

Thus, for many platforms more is not always better. The greater the potential for market failure, the greater the value that a platform sponsor can realize for all participants by imposing rules that limit or exclude those contributions that would diminish the overall value of the platform. In fact, most intra-platform

FOR MANY PLATFORMS
MORE IS NOT ALWAYS BETTER.

restraints likely fall into that category. Such intra-platform restraints are not designed to exclude would-be contributors that happen to also compete with the platform sponsor. Rather, the point of platform rules is to “help coordinate other players [who contribute to the platform] to achieve a better outcome than would be achieved in ungoverned production.”⁵⁴ In other words, the purpose of these intra-platform restraints is to make the platform more competitive vis-à-vis other platforms.⁵⁵

VI. Taxonomy of Common Platform Rules

In the computer platform context, there are common categories of intra-platform restraints with inter-platform effects, including rules about minimum quality, security, privacy, consistency, and technology.

A. QUALITY AND CONTENT RULES

Intra-platform rules ensuring quality are ubiquitous. In the gaming console space, Microsoft, Sony, and Nintendo all require developers to complete a multi-step process from concept approval to final testing to ensure the quality of a game prior to its release.⁵⁶ The problem that the sponsors seek to solve is that low-quality contributors do not fully internalize the costs that they impose on the more committed platform participants and might therefore have incentives to release poor products, turn a quick profit, and have other platform constituents suffer the consequences. Quality control has thus long been recognized as a bona fide business justification for vertical restraints and refusals to deal.⁵⁷ In the franchise context, franchisors are free to impose detailed “brand standards” on a franchisee’s business operations, even though such restraints clearly limit the dimensions of competition among franchisees and possibly between franchisees and the franchisor as well.⁵⁸

Content restraints are special cases of intra-platform quality rules. For example, Google’s sponsored search network AdWords has extensive “Content Guidelines” that prohibit advertisers from promoting liquor, tobacco, “the promotion of revisionist concepts,” etc.⁵⁹ Apple’s App Store Guidelines impose similar restrictions on iOS developers, rejecting applications that “portray[] realistic images of people or animals being killed or maimed, shot, stabbed, tortured or injured;” that “depict violence or abuse of children;” or that contain pornographic material.⁶⁰ Those rules have at times been singled out for criticism.⁶¹ However, the basis for such content restrictions—a newspaper would call them editorial

policies—is not to avoid competition but to create and maintain a distinctive platform identity or brand. Legislating and enforcing content rules are attempts by the platform sponsor to balance the interests of various constituencies with the goal of maximizing the overall value of the platform to better compete against rival platforms.

B. SECURITY RULES

Malware, i.e. malicious code that is executed on a machine unbeknownst to its user, is among the most serious threats to any computer platform.⁶² Malware can be used to steal a user's banking information and passwords or to press a user's PC into the service of a botnet to distribute spam or participate in distributed denial of service attacks. The first major malware incident on a mobile platform was reported in March 2011, when Google confirmed, “that 58 malicious applications were uploaded to Android Market, and that they were downloaded onto around 260,000 devices.”⁶³

Platform sponsors have long struggled with the problem of online security. Some are promoting *ex post* solutions, e.g., Microsoft offers a virus- and malware scanner as a free download.⁶⁴ Going one step further, Google's Android operating system includes a “kill signal,” to remotely remove infected applications from users' devices.⁶⁵ The problem with all *ex post* solutions is, however, that they cannot, by definition, protect against new threats, and that removing an infected application does not reliably disinfect already compromised devices. Apple has taken a different approach to improving security on iOS devices by screening all applications for malware prior to releasing them for download into the App Store.

RELATIVE PLATFORM SECURITY
IS AN IMPORTANT COMPETITIVE
DIFFERENTIATOR.

Relative platform security is an important competitive differentiator, as evidenced by scores of news articles and blog posts comparing the relative benefits of Android's *ex post* approach to the security benefits of Apple's *ex ante* approach, in the wake of the Android malware incident.⁶⁶ Such differentiation is once again realized through intra-platform developer restraints⁶⁷ and, in the case of Apple, the vertical integration of the App Store into the iOS. Viewing intra-platform security rules as restraints on developers only, without considering the effects on other platform participants and on inter-platform competition, would thus result in an incomplete and misleading assessment of their competitive effects.

C. PRIVACY RULES

Online privacy policies and privacy-respecting architectures have increasingly become a factor of competitive differentiation as well. In September 2008, Google cut back the retention period for IP addresses on server logs from 18 months to 9 months “to protect user privacy.”⁶⁸ Yahoo! responded quickly by cut-

ting back its retention time from 13 months to 90 days.⁶⁹ More recently, user privacy concerns—backed by the threat of regulatory intervention—led to numerous changes in Facebook’s privacy policies.⁷⁰ DuckDuckGo, an Internet search engine, has made user privacy its tagline (“We don’t track you!”) in an attempt to differentiate its service from Google.⁷¹ Similarly, FTC Commissioner Thomas Rosch has identified the need to balance consumer privacy and vigorous competition “as the most significant issue on the horizon in the high-tech sphere.”⁷²

Platforms compete along the privacy dimension through (a) privacy policies between the sponsor and the users and (b) sponsor-imposed intra-system restraints on developers and content publishers. These agreements and restraints seek to balance the privacy interests of users against the monetization interests of developers and advertisers. Clearly, such intra-platform rules translate into inter-platform competition for users and developers.⁷³

D. TECHNOLOGY RULES

Platform sponsors must determine what technologies to include in their platforms. For example, Microsoft had to decide what graphics library to include in the Xbox. That choice is made with both the contributors and the users in mind. The platform technologies must be able to deliver a compelling experience to the users but, in order to do so, they must first win the support of the developers. In the case of the Xbox, Microsoft chose to build the platform around its own DirectX graphics engine rather than OpenGL, a rival graphics library with broad developer support.

Can such choices be exclusionary? In the case of a “startup” platform, the answer will almost always be no. If there is no market power in the platform market, then the sponsor should be free to integrate whatever technologies it pleases, including its own. Now suppose that the platform becomes highly successful and achieves monopoly power. Would a requirement to keep using the sponsor-provided platform technologies at the expense of rival technologies be exclusionary?

IF THERE IS NO MARKET POWER IN
THE PLATFORM MARKET,
THEN THE SPONSOR SHOULD BE
FREE TO INTEGRATE WHATEVER
TECHNOLOGIES IT PLEASES,
INCLUDING ITS OWN.

In most instances, the answer will be no as well. Even for those technology mandates that could amount to exclusionary conduct under Section 2, there will often be compelling business justifications for a platform sponsor to manage the overall technology path by limiting alternative choices by developers and other contributors. Consistency of the user experience across applications can be a compelling reason, in particular if the platform vendor introduces a new, distinctive interface. So are privacy and security concerns, as discussed above. Similarly, maintaining a stable, predictable platform core simplifies application development, as developers do not have to customize their software for different hardware configurations.⁷⁴

Other technology restrictions—e.g., native development-only requirements—that prohibit developers from using cross-platform intermediation layers may also be justified. Intermediation layers sit between the platform’s operating system APIs and the applications. Instead of writing applications directly for a given operating system, developers write applications for the intermediation layer, which then connects to the operating system. The main benefit of cross-platform intermediation is that developers can run the same code on multiple platforms. Without cross-platform intermediation, developers would have to “port” their programs from one platform to another, which—depending on the program—can be rather time and labor intensive.⁷⁵ Why would a platform sponsor want to limit the use of such intermediation layers?

From a platform sponsor’s perspective, there are two related concerns.⁷⁶ First, widespread use of an intermediation layer by developers may require the platform sponsor to delay operating system or hardware innovation until the vendor has made the intermediation layer compatible with the upgrade. Second, many new operating system or hardware features may not be adopted by developers and thus remain invisible to users until the vendor chooses to expose the new features to developers in the intermediation layer.

These concerns are particularly serious for smaller and/or highly innovative platforms. Suppose that an intermediation layer runs on platforms A, B, and C. Platform C accounts for 5 percent of the layer’s installations and is innovating rapidly. In that case, the vendor’s incentives to upgrade the intermediation layer in lockstep with C is significantly weaker than the incentives of native platform C developers. The intermediation layer vendor’s incentive is to maintain maximum compatibility across platforms, not maximum quality on any given platform.

Sponsors of platforms whose success depends on rapid innovation may thus have strong business justifications for requiring developers to write native applications and not use intermediation layers, as rapid platform innovation and close-in-time exposure of new features to users are key attributes of inter-platform competition.⁷⁷

SPONSORS OF PLATFORMS WHOSE SUCCESS DEPENDS ON RAPID INNOVATION MAY THUS HAVE STRONG BUSINESS JUSTIFICATIONS FOR REQUIRING DEVELOPERS TO WRITE NATIVE APPLICATIONS AND NOT USE INTERMEDIATION LAYERS.

VII. Calibrating Antitrust Policy in the Systems Competition Context

Our initial question whether “closed” systems are inherently anticompetitive can be restated as follows: “Is there a reason to believe that intra-platform restraints imposed by the platform sponsor on various contributors are commonly exclu-

sionary?” To that question, the answer is no. Is it possible that such restraints can lead to anticompetitive exclusion? Yes, but not unless the platform has significant market power vis-à-vis rival platforms.

A. AS A POLICY MATTER, “OPEN” IS NOT NECESSARILY BETTER THAN “CLOSED”

From a competition policy point of view, the assumption that “open” is necessarily preferable to “closed,” requires some qualification.

First, we should not lose sight of the fact that the two great information monopolies of the Internet Age (as of Spring 2011), Microsoft and Google, have both been sponsors of open systems. Both firms succeeded in dominating broad horizontal layers of their respective stacks: the PC and the internet. Microsoft Windows, protected by the applications barrier to entry that its open developer access policies enabled, monopolized the operating system layer.⁷⁸ Google has become the “current custodian of the [internet’s] Master Switch”⁷⁹ as the dominant provider of algorithmic and sponsored search.⁸⁰ Less “open” platforms such as Facebook, Apple, and Twitter have remained in much more narrow verticals that dissatisfied customers and competitors can more easily avoid.

Second, while an “open monopolist” will generally be preferable to a “closed monopolist” from an antitrust policy standpoint, it is by no means clear that open systems are preferable to closed systems in competitive markets. For example, in the 1980s, before the Windows/Intel systems came to dominate the PC

SECOND, WHILE AN “OPEN MONOPOLIST” WILL GENERALLY BE PREFERABLE TO A “CLOSED MONOPOLIST” FROM AN ANTITRUST POLICY STANDPOINT, IT IS BY NO MEANS CLEAR THAT OPEN SYSTEMS ARE PREFERABLE TO CLOSED SYSTEMS IN COMPETITIVE MARKETS.

space, there was intense competition among vertically integrated home computing firms, including Sinclair, Altair, Tandy, Commodore, Atari, Texas Instruments, Sharp, Apple, and many more, each of which sold “the whole widget.” The competing systems were incompatible, which limited their value to users and developers alike.

However, for the same reason, innovation by the platform sponsor was not limited to a single layer of the system architecture. Each firm competed on the value of the entire ecosystem. The rapid pace of innovation that users and developers enjoyed in this golden age of closed computer systems competition remains somewhat underappreciated in today’s discussion. Yes, the open Windows/Intel PC blew away the competing walled gardens in much the same manner that the open internet blew away CompuServe, Prodigy, and AOL. But the new “open” era, for all its undeniable benefits, also ushered in decades of monoculture—replacing a wildly diverse collection of ecosystems, teeming with radical innovative experiments at all levels, with relative heterogeneity.

B. MONOPOLY POWER IS A THRESHOLD ISSUE FOR ANY CONCERNS OVER “CLOSED” SYSTEMS

Intra-system restraints are most comparable to vertical restraints such as franchise regulations and brand standards. Some rules may also resemble selective refusals to deal and cases about technological innovation that breaks or limits compatibility with third-party contributions. All these doctrinal categories share a market power requirement, either as part of a Section 1 rule of reason inquiry or as part of a Section 2 monopolization case. Without significant market power and the threat of a high level of market foreclosure, there is no basis for imposing antitrust liability on a platform sponsor’s “intra-platform legislation.”

This fundamental point does not always receive proper attention. For example, in *The Master Switch*, Wu proposes a “Separations Principle” to keep closed systems from taking over the information sector.

“[A Separations Principle] would mean that those who develop information, those who own the network infrastructure on which it travels, and those who control the tools or venues of access must be kept apart from one another. . . . [The resulting priorities for antitrust enforcement] must be both the prevention and dissolution of large-scale vertical mergers in the communications industry.⁸¹”

The Separations Principle amounts to a general rule against vertical integration in the information sector irrespective of market power, foreclosure, and efficiencies. Such a sweeping rule requires extraordinarily strong justifications, which Wu fails to provide. In fact, our analysis of the competitive effects of open and closed systems does not suggest that closed systems pose anywhere near the level of concern that would justify such a radical expansion of antitrust market regulation.⁸²

C. THERE ARE NO SHORTCUTS FOR A FINDING OF MONOPOLY POWER

Users and developers engage with platforms over time. They join a platform with certain expectations about its evolution, make follow-on investments in time, money, and expertise, periodically recommit (e.g., when a new device is rolled out or an operating system gets a major version release), or move on to other platforms.

Plaintiffs have taken snapshots of this fluid multi-party platform relationship and recast it as a simple two-step purchasing pattern. First, or so the argument goes, users and developers choose among available platforms (inter-platform

choice). Second, following the platform commitment, they choose from among the options available on the platform (intra-platform choice). The choice sets for each step may be very different. For example, a developer seeking to enter the mobile application business may choose among Android, Apple, RIM, Windows Mobile, WebOS, and other platforms. Clearly, there is abundant choice at the inter-platform level. Once the developer has committed to a platform, however, its options within that platform may be limited by platform rules, technologies, the platform’s user base, etc.⁸³

Under this snapshot approach, the two-step purchasing pattern then serves as the basis for applying a *Kodak*-style single-product aftermarket theory.⁸⁴ The platform is the foremarket, and everything else is part of various aftermarkets. To the extent that the platform sponsor participates in those aftermarkets and reserves certain business opportunities for itself (e.g., by regulating the behavior of other contributors or various forms of technological integration), it has been cast as a putative monopolist.⁸⁵

A closer examination reveals that any similarities between *Kodak*-style aftermarkets and computer platforms are superficial and uninformative.

First, while “Kodak’s sale of its product involved no contract framework for ongoing relations,”⁸⁶ users and developers enter into long-term relationships with the platform sponsor. Unlike the buyer of a copier, users and developers join evolving ecosystems with potentially tens of thousands of contributors and a “platform government” whose job it is to—more or less—actively manage the overall platform path by constantly fine-tuning the rules governing the relationship among the platform constituents. In other words, if it is reduced to a series of spot transactions, then something important about the long-term nature of platform engagement is lost.

A CLOSER EXAMINATION REVEALS
THAT ANY SIMILARITIES BETWEEN
KODAK-STYLE AFTERMARKETS AND
COMPUTER PLATFORMS ARE
SUPERFICIAL AND UNINFORMATIVE.

Second, and relatedly, buyers of copiers may have certain expectations of stability, both in terms of the functionality of the product and its general service environment. In contrast, users, developers, and platform sponsors all expect change, even with respect to core functionalities of the devices that define the platform. Copiers don’t fundamentally change over the course of their useful life, but computers do. After all, as Zittrain correctly observes, computers are generative devices. The same is true with respect to the rules governing use and contributions. End user license agreements are amended frequently and so are developer terms as the ecosystem adapts to internal and external challenges.⁸⁷

Third, the *Kodak* court was concerned about “a less responsive connection” between the equipment market and the aftermarkets for parts and services as a result of information and switching costs.⁸⁸ In other words, *ex ante* buyer ignorance about the value of third-party aftermarket contributions creates the opportunity

for *ex post* exploitation. Because of that, or so the theory goes, equipment vendors have no incentives to educate the buyer at the time of the equipment purchase.

The incentives of most computer platform sponsors are completely different. User ignorance about the value of third-party contributions is a serious problem, not a business opportunity. As a result, platform sponsors aggressively advertise developer contributions. The iPhone is more attractive because “[t]here is an app for that.” Android is more attractive because “Droid does apps.” Microsoft advertises third party games for the Xbox. Sony and Nintendo do the same. Platform sponsors that are not merely selling “a device” but the value of the entire ecosystem have strong incentives to create highly responsive connections between the platform and third party contributions—to the point that a temporal distinction between the two (“first platform, then third-party contribution”) becomes artificial at best and misleading at worst. “Joining a platform” means entering a web of evolving, long-term relationships that jointly create the competitive value of a platform. It is that value that determines the initial platform choices of users and developers and that underwrites their continued support.

AFTERMARKET THEORIES
ARE LIKELY TO MISCHARACTERIZE
THE COMPETITIVE ENVIRONMENT
OF COMPUTER PLATFORMS
AND SUGGEST MONOPOLY
PROBLEMS WHERE THERE
ARE NONE.

As a result, aftermarket theories are likely to mischaracterize the competitive environment of computer platforms and suggest monopoly problems where there are none.⁸⁹ Courts have reacted to a similar “false positive” problem in the context of franchise agreements by rejecting aftermarket theories where the source of the franchisor’s power is based on an agreement.⁹⁰ Computer systems deserve at least the same level of deference. There are no shortcuts to a finding of real, as opposed to imputed, inter-platform market power.

D. ADVERSE EFFECTS OF INTRA-PLATFORM RESTRAINTS ON ONE CONSTITUENCY MUST BE BALANCED AGAINST EFFECTS ON OTHER CONSTITUENCIES AND AGAINST INTER-PLATFORM EFFECTS

The antitrust evaluation of open and closed platforms should focus on real (not aftermarket-imputed) market power in the inter-platform space and the net competitive effect of the intra-platform rule(s) under consideration.

- (1) If there is no meaningful inter-platform market power, then regulatory intervention is unwarranted. There is no reason to view intra-platform rules less favorably than other vertical intra-brand restraints.
- (2) If there is significant inter-platform market power, then any meaningful competitive effects from the intra-platform rule *vis-à-vis* the restrained platform constituency should be balanced against (a) benefits that the rule confers upon other platforms constituencies and (b) its positive effects (if any) on inter-platform rivalry.⁹¹

At bottom, the bad reputation of closed systems or walled gardens in the “open versus closed” debate is quite undeserved. Walled gardens generally benefit their environments—both in the real world and the digital realm. The primary purpose of a garden wall, after all, is to shelter plants from wind and frost, not to keep intruders out.⁹² In the protected space of the garden, flowers can grow that would not otherwise survive in the wild. Walled gardens thus deliberately create a microcosm that is different from the surrounding ecosystem. Therefore, as long as the garden does not take over the entire ecosystem, walled gardens increase, not reduce, overall diversity. From a competition policy perspective, enjoying the fruits of a walled garden is generally not a guilty pleasure. ▼

-
- 1 Jonathan Zittrain, *The Future of the Internet And How To Stop It* (2008) (“Future”), available at <http://futureoftheinternet.org/>.
 - 2 Ed Felten, *iPad to Test Zittrain’s ‘Future of the Internet’ Thesis, Freedom To Tinker* (February 4, 2010), available at <http://www.freedom-to-tinker.com/blog/felten/ipad-test-zittrains-future-internet-thesis>
 - 3 Zittrain, *Future*, *supra* note 1 at 3.
 - 4 *Id.* at 90.
 - 5 TIM WU, *THE MASTER SWITCH: THE RISE AND FALL OF INFORMATION EMPIRES*, p. 6 (2010) (“The Master Switch”).
 - 6 *Id.* at 296.
 - 7 *Id.* at 296 - 97. Another civic virtue of generative systems is their greater resilience against government control, because there is no single point of failure, no single door that the governments of the world could knock on to silence unwelcome speech. Zittrain, *A Fight over Freedom at Apple’s Core*, *FINANCIAL TIMES* (February 3, 2010), available at <http://www.law.harvard.edu/news/2010/02/04-zittrain-apple.html>. The recent experiences in China, Egypt, Iran, Libya, and Tunisia, and the U.S. government’s response to the release of U.S. diplomatic cables by Wikileaks have brought this particular aspect of the “open versus closed” debate into the spotlight. However, as Adam Thierer persuasively argues, “[i]nstead of imposing restrictions on code or coders to limit regulability, we should instead place more constraints on our government(s).” Adam Thierer, *The Case for Internet Optimism, Part 2: Saving the Net from its Supporters*, in: BERIN SZOKA & ADAM MARCUS, *THE NEXT DIGITAL DECADE*, p. 148 (2010), available at http://nextdigitaldecade.com/ndd_book.pdf. In any event, the close connection to civic values and underlying political philosophies is no doubt a reason for the occasional rhetorical overheating of the debate.
 - 8 Zittrain, *Future*, *supra* note 1 at 79.
 - 9 *Id.*, at 89.
 - 10 “[T]here’s a lot of stuff I would do for free that I wouldn’t do for a pittance.” Cory Doctorow (<http://www.craphound.com>), quoted in: Matt Cowan & Georgina Proshan, *Wikipedia founder sees biggest impact yet to come*, *REUTERS* (January 11, 2011), available at <http://www.reuters.com/article/2011/01/14/us-wikipedia-wales-idUSTRE70D31X20110114>.
 - 11 Jonathan Zittrain, *Protecting the Internet Without Wrecking It: How to Meet the Security Threat*, in: BERIN SZOKA & ADAM MARCUS, *THE NEXT DIGITAL DECADE*, p. 98 (2010).

- 12 Tim Berners-Lee, *Long Live the Web: A Call for Continued Open Standards and Neutrality*, SCIENTIFIC AMER. (November 22, 2010), available at <http://www.scientificamerican.com/article.cfm?id=long-live-the-web>.
- 13 Jonathan Rosenberg, *The Meaning of Open*, Official Google Blog (December 21, 2009), available at <http://googleblog.blogspot.com/2009/12/meaning-of-open.html>. That view is not universally shared. See, e.g., Brian Prentice, *The Truth of Open*, Gartner Blog Network (December 22, 2009), available at http://blogs.gartner.com/brian_prentice/2009/12/22/the-truth-of-open/. ("The truth is that closed systems still win. Open systems, practically speaking, are basically good for making others lose.")
- 14 Tim Wu, quoted in: Nick Bilton, *One on One: Tim Wu, Author of The Master Switch*, N.Y. TIMES (November 14, 2010), available at <http://bits.blogs.nytimes.com/2010/11/14/one-on-one-tim-wu-author-of-the-master-switch/>.
- 15 Tim Wu, *How Apple's Closed Ways Could Land It Into Antitrust Trouble*, TECHCRUNCH (November 20, 2010), available at <http://techcrunch.com/2010/11/20/apple-antitrust/>.
- 16 *Id.*
- 17 Tom Conlon, *The iPad's Closed System: Sometimes I Hate Being Right*, popsci.com (Jan. 29, 2010), available at <http://www.popsci.com/gadgets/article/2010-01/ipad's-closed-system-sometimes-i-hate-being-right>. The "open v. closed" debate is not limited to computer systems. In the context of the proposed acquisition of NYSE Euronext by Deutsche Börse AG, Commissioner Joaquin Almunia stated: "I tend to prefer models that are not a vertical silo. . . . This creates more competition, more opportunities and this [i.e., NYSE's] more open business model—together with interoperability—from the competition point of view is the preferred one." Aofi White, *Deutsche Boerse, NYSE Deal Faces Lengthy EU Antitrust Review, Almunia Says*, Bloomberg.com (March 22, 2011), available at <http://www.bloomberg.com/news/2011-03-22/deutsche-boerse-nyse-deal-faces-lengthy-eu-review-competition-chief-says.html>.
- 18 The civic and political aspects of the open versus closed debate are beyond the scope of our discussion. Of course, a neat separation of economic and civic matters is not always possible, but focusing on one aspect of a phenomenon does not deny the existence or the significance of other aspects. One can, without contradiction, support open systems as a matter of civic commitment and, at the same time, conclude that antitrust regulation is not the right tool to promote their adoption.
- 19 Michael Katz & Carl Shapiro, *Systems Competition and Network Effects*, 8 J. ECON. PERS. 93 (1994), available at http://socrates.berkeley.edu/~scotch/katz_shapiro.pdf.
- 20 NIKLAS LUHMANN, SOCIAL SYSTEMS (1996); G. SPENCER-BROWN, LAWS OF FORM (1969).
- 21 Zittrain, Future, *supra* note 1 at 16-17.
- 22 Wu, *The Master Switch*, *supra* note 5 at 279 ("The victory of PCs and Windows over Apple was viewed by many as the defining parable of the decade; its moral was 'open beats closed.'").
- 23 See, e.g., John Sullivan, *Free software is a matter of liberty, not price*, Free Software Foundation (March 7, 2011), available at <http://www.fsf.org/about/> and *Windows 7 Sins* (last accessed March 5, 2011), available at: <http://en.windows7sins.org/>.
- 24 Zittrain, Future, *supra* note 1 at 16 ("Windows PCs, like the Mac OS and Linux counterparts . . . welcome code from any source.").
- 25 Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age*, 17 HARV. J. LAW & TECH. 85, 92

- (2003), available at <http://ssrn.com/abstract=452220> or doi:10.2139/ssrn.452220 ("Apple [with the 1984 Macintosh] relied on a closed business model."). Wu, *The Master Switch*, *supra* note 5 at 276-78 (Steve Jobs' "decision to close the Macintosh contributed significantly to making Bill Gates the richest man on earth.").
- 26 See, e.g., Zittrain, *Future*, *supra* note 1 (Locked-down devices include "iPods, most video game consoles, e-book readers like the Amazon Kindle, and cable-company set-top boxes); at 64 (TiVo as "information appliance"); at 77 ("[I]t is nearly impossible for the Linux PC inside a TiVo to run anything but the code that TiVo designates for it"); Joe Brockmeier, *Amazon, Open the Kindle before Apple eats your lunch*, ZDNET (October 2, 2008), available at <http://www.zdnet.com/blog/community/amazon-open-the-kindle-before-apple-eats-your-lunch/114> ("Since the Kindle is a closed platform, Amazon misses out on driving sales through add-on applications that could add value to its device.")
- 27 Zittrain, *Future*, *supra* note 1 at 2-3. In February 2008, Apple opened the iPhone to third-party software developers.
- 28 Zittrain, *A Fight over Freedom at Apple's Core*, *supra* note 7 ("The iPhone's hybrid model of centrally controlled outside software is already moving beyond the smart phone.").
- 29 See Wikipedia entry for "App Store," available at http://en.wikipedia.org/wiki/App_Store (last accessed March 10, 2011).
- 30 See e.g., Joseph Farrell & Philip J. Weiser, *supra* note 25, at 87 (2003); HERBERT HOVENKAMP, *THE ANTITRUST ENTERPRISE: PRINCIPLE AND EXECUTION*, pp. 182, 207 (2005).
- 31 Joseph Farrell, Hunter K. Monroe, & Garth Saloner, *The Vertical Organization of Industry: Systems Competition versus Component Competition*, 7 *J. ECON. & MANAG. STRAT.* 143, 144 (1998).
- 32 *Id.*
- 33 *Brown Shoe v. United States*, 370 U.S. 294 (1962).
- 34 *U.S. v. Arnold, Schwinn & Co.*, 388 U.S. 365, 379 (1967).
- 35 Joseph Farrell & Philip J. Weiser, *supra* note 25 at 95 ("Modular industry structures enable independent firms to introduce innovations into an established environment. An open architecture can facilitate innovation in individual components, spur market entry, and result in lower prices.").
- 36 *Continental T.V. v. GTE Sylvania*, 433 U.S. 36 (1977); *Business Electronic Corp. v. Sharp Electronics Corp.*, 485 U.S. 717, 725 (1988).
- 37 William E. Kovacic & Carl Shapiro, *Antitrust Policy: A Century of Economic and Legal Thinking*, 14 *J. ECON. PERSPECTIVES* 43, 53-54 (2000), available at <http://faculty.haas.berkeley.edu/shapiro/century.pdf>.
- 38 *Leegin Creative Leather Products, Inc. v. PSKS, Inc.*, 551 U.S. 877 (2007). Doctrinally, tying, a vertical inter-brand restraint, still exists as a per se offense, hence the hedge ("almost universally"). However, tying is per se illegal only in name, as per se tying first requires proof that the supplier possesses significant market power over the tying product. That, of course, renders the per se label all but meaningless. The point of a per se offense is that the conduct is illegal irrespective of market power. See generally, AREEDA & HOVENKAMP, *ANTITRUST LAW* ¶1702 (2011 Online Edition).
- 39 Quoted in Chris Ziegler, *Nokia CEO Stephen Elop rallies troops in brutally honest 'burning platform' memo?*, ENGADGET (February 8, 2011), available at <http://www.engadget.com/2011/02/08/nokia-ceo-stephen-elop-rallies-troops-in-brutally-honest-burnin/>.

- 40 See Thomas Eisenmann, Geoffrey Parker, & Marshall Van Alstyne, *Opening Platforms: How, When, Why?* 3 (2008 Working Paper), *available* at <http://ssrn.com/abstract=1264012>.
- 41 The sponsor may be a single firm, as is usually the case with proprietary platforms, an organization (e.g., the Mozilla Foundation), or a loosely organized group of contributors (e.g., Linux).
- 42 For examples from console game development, see DAVID EVANS, ANDREI HAGIU, & RICHARD SCHMALENSEE, *INVISIBLE ENGINES: HOW SOFTWARE PLATFORMS DRIVE INNOVATION AND TRANSFORM INDUSTRIES*, p. 130 (2006) (“Invisible Engines”).
- 43 The availability of (killer-) application X on platform A, however, may well be an important element of inter-platform competition. E.g., every mobile computing platform is going out of its way to advertise the availability of Facebook and Twitter clients.
- 44 See www.theladders.com (last accessed, March 10, 2011). The vast majority of the literally hundreds of competing platforms only charge the employers.
- 45 See, e.g., Microsoft, “Developing for Xbox LIVE Arcade,” *available* at <http://www.xbox.com/en-US/Community/Developer/DevelopingForLiveArcade> (last accessed, March 5, 2011).
- 46 Jonathan Rosenberg, quoted in: Google, *Inside the Black Box Technology and Innovation at Google* (February 27, 2008), *available* at <http://bit.ly/jrosenberg2008>.
- 47 Case No COMP/M.5727—Microsoft/Yahoo! Search Business (2010), ¶¶ 47-48.
- 48 Kevin J. Boudreau & Andrei Hagiu, *Platform Rules: Multi-sided Platforms as Regulators*, Working Paper 3, 5 (2008), *available* at <http://ssrn.com/abstract=1269966>. (“Hypothesis 1: Markets around [Multi-Sided Platforms] are riddled with externalities and other sources of coordination problems.”).
- 49 Evans, Hagiu, & Schmalensee, *INVISIBLE ENGINES*, *supra* note 42 at 3.
- 50 *U.S. v. Microsoft*, 253 F.3d 34, 55 (D. C. Cir. 2001).
- 51 A same-side effect of crowding, akin to a pecuniary externality, is “too much competition” among developers, making competing platforms with fewer users relatively more attractive. As a result, there may be same-side intra-platform limits to growth that a platform sponsor may seek to minimize, e.g., by limiting the number of applications in a given category.
- 52 Significant enough to provide a business opportunity for application discovery services such as Chomp, *available* at <http://www.chomp.com>.
- 53 See, e.g., RICHARD IPPOLITO, *ECONOMICS FOR LAWYERS*, p. 284 (2005). This particular type of market failure in which bad products drive out good products as a result of information asymmetries is sometimes referred to as the “lemons problem.”
- 54 Boudreau & Hagiu, *Platform Rules*, *supra* note 48 at 5.
- 55 If nightclub A imposes intra-platform restraints on contributions by turning away disorderly patrons, it does so to compete more effectively against other nightclubs. If nightclub A gains a reputation for “safety,” then nightclub B will either have to impose similar rules (“we’re just as safe if not safer”), differentiate its offering through more lenient rules (“everyone’s welcome”), or differentiate its offering through rules emphasizing different aspects of the nightclub experience (“A may be safer, but B is more fun”).

- 56 Evans, Hagiü, & Schmalensee, *Invisible Engines*, *supra* note 42 at 125-136. The specter of Atari's downfall in 1982-83, caused in part by low-quality third party cartridges and the poor performance of the Pac-Man and E.T. titles on the VCS 2600, still haunts the gaming industry. STEVEN KENT, *THE ULTIMATE HISTORY OF VIDEO GAMES*, pp. 236, 238-39 (2001). See also Boudreau & Hagiü, *Platform Rules*, *supra* note 48 at 1.
- 57 See, e.g., *U.S. v. Microsoft Corp.*, 253 F.3d 34, 59 (D.C. Cir. 2001); *Aspen Highlands Skiing Corp. v. Aspen Skiing Co.*, 472 U.S. 585, 608-12 (1985).
- 58 See generally, HOVENKAMP, *THE ANTITRUST ENTERPRISE*, pp. 204-6 (2005).
- 59 See Google's Advertising Policies Content Guidelines, *available* at <http://adwords.google.com/support/aw/bin/static.py?hl=en&topic=28436&guide=28435&page=guide.cs> (last accessed March 3, 2011).
- 60 Apple App Store Review Guidelines at §§ 15-16, 18-19.
- 61 See, e.g., Tim Bray, *Now a No-Evil Zone*, tbray.org (Mar. 15, 2010) ("The iPhone vision of the mobile internet's future omits controversy, sex, and freedom, but includes strict limits on who can know what and who can say what. . . . I hate it."), *available* at <http://www.tbray.org/ongoing/When/201x/2010/03/15/Joining-Google>.
- 62 Zittrain, *Future*, *supra* note 1 at 36-61.
- 63 Jason Kincaid, *Google Responds To Android Malware, Will Fix Infected Devices And 'Remote Kill' Malicious Apps*, *TECHCRUNCH* (March 5, 2011), *available* at: <http://techcrunch.com/2011/03/05/android-malware-rootkit-google-response/>.
- 64 <http://www.microsoft.com/security/pc-security/windows7.aspx>.
- 65 Jason Kincaid, *Google Responds*, *supra* note 63.
- 66 See, e.g., Dan Goodin, *Android Malware attacks show perils of Google openness*, *THE REGISTER* (March 4, 2011), *available* at http://www.theregister.co.uk/2011/03/04/google_android_market_peril/; Joseph Menn, *Google Disables Android Malware*, *FINANCIAL TIMES* (March 7, 2011), *available* at: <http://www.ft.com/cms/s/0/f04a88b8-48ea-11e0-af8c-00144feab49a.html#axzz1GcEHOURW>. As to online payments for app purchases, convenience and security are the users' primary concerns, see *THE NIELSEN COMPANY*, *THE STATE OF MOBILE APPS 5* (2010), *available* at <http://blog.nielsen.com/nielsenwire/wp-content/uploads/2010/09/NielsenMobileAppsWhitepaper.pdf>.
- 67 See, e.g., *Apple App Store Review Guidelines*, §2.4 ("Apps that include undocumented or hidden features inconsistent with the description of the app will be rejected").
- 68 Peter Fleischer, Kane Horvath, & Alma Whitten, *Another step to protect user privacy*, *Official Google Blog* (August 9, 2008), *available* at <http://googleblog.blogspot.com/2008/09/another-step-to-protect-user-privacy.html>.
- 69 Anne Toth, *Your data goes incognito*, *YODEL ANECDOTAL* (December 17, 2008), *available* at <http://ycorpblog.com/2008/12/17/your-data-goes-incognito/>
- 70 Facebook's privacy policy is *available* at <http://www.facebook.com/policy.php> (last visited March 12, 2011). For a summary of reactions to recent changes see e.g., Brennon Slattery, *Facebook's Privacy Changes Get Mixed Reviews*, *PC WORLD* (May 27, 2010), *available* at http://www.pcmag.com/article/197359/facebook_privacy_changes_get_mixed_reviews.html.
- 71 <http://duckduckgo.com/privacy.html> (last accessed, March 1, 2011).

- 72 J. Thomas Rosch, Intel, Apple, Microsoft, and Facebook: Observations on Antitrust and the High-Tech Sector 20 (November 18, 2010), *available* at <http://ftc.gov/speeches/rosch/101118fallforum.pdf>.
- 73 Privacy restraints also reflect differences in the relative bargaining power of platform constituencies across platforms. For example, Apple's relatively strong commitment to user privacy is undoubtedly influenced by the fact that payments from users account for 93 percent of Apple's revenues. (See Apple, Inc. 2010 Annual Report, filed Sep. 29, 2010 on Form 10-K, p. 33.) By the same token, Google's relatively weak privacy rules reflect the business reality that 97 percent of Google's revenues come from advertisers. (Google, Inc. 2009 Annual Report, filed Feb. 12, 2010 on Form 10-K, p. 37).
- 74 Evans, Hagiu, & Schmalensee, *INVISIBLE ENGINES*, *supra* note 42 at 197, 331 ("[S]tandardizing software platforms tends to help both users and applications developers").
- 75 Much depends on the developer's skill and the project design. All else being equal, native development will yield better performance. Thus, many developers who target multiple platforms design their projects for easy portability from the ground up, which is considered good programming practice. For example, John Carmack, the legendary id Software programming lead, ported Quake to OpenGL by himself in a weekend. DAVID KUSHNER, *MASTERS OF DOOM*, p. 227 (2004).
- 76 Probably the best-known example of a platform sponsor attempting to limit the use of an intermediation layer involved Microsoft's attacks on Java in the mid-1990s. According to the D.C. Circuit, Microsoft feared that Java, combined with the Netscape browser, could erode the applications barrier to entry and, further down the road, mature into an alternative to the Windows operating system itself. Some of Microsoft's anti-Java actions were thus found to constitute unlawful monopoly defense under Section 2. With the benefit of hindsight, we now know that this concern was most likely overblown. Moreover, the Microsoft case is somewhat atypical, as intermediation layers are usually more limited than Java and even less likely to evolve into replacements of full-fledged operating systems. *U.S. v. Microsoft*, 253 F.3d 34, 74-77 (D. C. Cir. 2001).
- 77 In this context, see Michael Katz & Carl Shapiro, *Systems Competition*, *supra* note 19 at 95 ("It is tempting, but misleading, to view incompatibility as just another coordination failure. Although compatibility has obvious benefits, obtaining and maintaining compatibility often involves a sacrifice in terms of product variety or restraints on innovation."); Joseph Farrell & Philip J. Weiser, *Modularity*, *supra* note 25 at 99 ("Innovation can require changing the platform/application interface, which can be a slow process if an industry relies on open standards and open interfaces. In such cases, hand-in-glove coordination between the platform sponsor and one or more complementors can accelerate innovation.")
- 78 *U.S. v. Microsoft*, 253 F.3d 34, 57 (D. C. Cir. 2001).
- 79 Wu, *The Master Switch*, *supra* note 5 at 279.
- 80 U.S. DOJ, Yahoo! Inc. and Google Inc. Abandon Their Advertising Agreement (November 5, 2008), *available* at <http://www.justice.gov/opa/pr/2008/November/08-at-981.html>.
- 81 Wu, *The Master Switch*, *supra* note 5 at 304, 311.
- 82 See also, Adam Thierer, *Thoughts on Wu's Master Switch, Part 6*, *TECHLIBERATION* (November 2, 2010), *available* at <http://techliberation.com/2010/11/02/thoughts-on-wu-s-master-switch-part-6-his-audacious-information-industrial-policy/>.
- 83 This stylized view ignores multi-homing by both users and developers, which is common on platforms where users value variety relatively more than compatibility, e.g., gaming consoles. See, e.g., Evans, Hagiu, & Schmalensee, *Invisible Engines*, *supra* note 42 at 139.

- 84 Eastman Kodak v. Image Technical Services, 504 U.S. 451, 486 (1992). Carl Shapiro, *Aftermarkets and Consumer Welfare: Making Sense of Kodak*, 63 ANTITRUST L.J. 486 (1995). For a critical assessment, see Daniel M. Wall, *Aftermarket Monopoly: Five Years After Kodak*, ANTITRUST 32 (1997).
- 85 See, e.g., Apple, Inc. v. Psystar Corp., 673 F. Supp. 2d 931, 1200 (N.D. Cal. 2009) ("Psystar alleges that, by tying or other exclusionary behavior, Apple seeks to dominate a distinct aftermarket, the market for Mac OS-compatible hardware systems."); Dattel Holdings v. Microsoft Corp., 712 F.Supp.2d 974, 980 (N.D. Cal. 2010). ("[A]ftermarket for Xbox 360 accessories.").
- 86 Alan Silberman, *The Myths of Franchise Market Power*, 65 ANTITRUST L.J. 217 (1996).
- 87 See, e.g., Apple, Inc. v. Psystar Corp., 586 F. Supp. 2d 1190, 1201 (N.D. Cal. 2008), where the court rejected a single-product aftermarket, because the EULA "specifically restricts the use of Mac OS to Apple-labeled computer hardware systems. Consumers, therefore, knowingly agree to the challenged restraint."
- 88 Eastman Kodak Co. v. Image Technical Services, 504 U.S. 451, 473 (1992) ("[T]he existence of significant information and switching costs . . . could create a less responsive connection between service and parts prices and equipment sales.")
- 89 Dan Wall, *supra* note 84 at 33 ("Aftermarkets . . . present something of an anomaly: a frequently recurring set of conditions that when assessed according to orthodox antitrust analysis point to what should be a rarity, namely monopoly.")
- 90 Queen City Pizza, Inc. v. Domino's Pizza, Inc., 124 F. 3d 430, 440 (3rd. Cir. 1997) ("[P]laintiffs here knew that Domino's Pizza retained significant power over their ability to purchase cheaper supplies from alternative sources because that authority was spelled out in detail in section 12.2 of the standard franchise agreement. Unlike the plaintiffs in Kodak, the Domino's franchisees could assess the potential costs and economic risks at the time they signed the franchise agreement. . . . Kodak's sale of its product involved no contractual framework for continuing relations with the purchaser. But a franchise agreement regulating supplies, inspections, and quality standards structures an ongoing relationship between franchisor and franchisee designed to maintain good will. These differences between the Kodak transaction and franchise transactions are compelling.")
- 91 In the context of a Section 2 inquiry, such intra-platform and inter-platform benefits are part of a platform sponsor's business justifications. In many cases, the mere existence of a non-pretexual business justification should suffice to render a monopolization claim implausible. In those instances, no elaborate balancing test is required. In appropriate cases, the reasoning in *Allied Orthopedic Appliances v. Tyco Healthcare Group*, 592 F.3d 991 (9th Cir. 2010) suggests even greater deference to the sponsor's decisions: "To weigh the benefits of an improved product design against the resulting injuries to competitors is not just unwise, it is unadministrable. There are no criteria that courts can use to calculate the 'right' amount of innovation, . . . A seemingly minor technological improvement today can lead to much greater advance in the future." *Id.*, 1000.
- 92 See Wikipedia entry for "Walled Garden," available at http://en.wikipedia.org/wiki/Walled_garden (last accessed, March 3, 2011).



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

The Emergence of Global Search Engines: Trends in History and Competition

Manish Agarwal & David K. Round

University of South Australia

The Emergence of Global Search Engines: Trends in History and Competition

*By Manish Agarwal & David K. Round**

The web search market is an example of a two-sided market where internet users account for one side and the advertisers for the other. Given the increased regulatory scrutiny faced by the web search market, this paper uses the two-sided market framework to analyze the market structure and the behavioral trends on both sides of the market in order to assess the state of competition in this market. Section 2 traces the evolution of the web search engines. Section 3 presents the two-sided market framework and examines trends on both sides of the web search market. Section 4 concludes.

*Manish Agarwal is Research Fellow, Centre for Regulation and Market Analysis, School of Commerce, University of South Australia and David K. Round is Professor of Economics, and Director Centre for Regulation and Market Analysis, School of Commerce, University of South Australia

I. Introduction

Competition agencies around the world are increasingly paying attention to the web search market with much of the regulatory scrutiny focused upon the search rankings and advertising practices of the dominant web search engines.

In December 2010, the French competition authority concluded a study examining competition in the online advertising sector in France. It observed that Google held a dominant position in the advertising market linked to search engines and identified possible exclusionary conduct and abuses that would merit further investigation.¹ In November 2010, the European Commission began an antitrust investigation into Google's online search and advertising practices as a result of complaints that Google discriminates against websites that offer competing online services.² There are news reports that the U.S. antitrust regulators may open an investigation into Google's dominance of the web search industry, an action that has been endorsed by some commentators.³ Echoing a similar case, Baidu, the popular Chinese web search engine, is facing a complaint that it manipulated its search results to block or lower the ranking of a Chinese online encyclopedia, Hudong.⁴

A common feature underlying these cases is that the web search engine provider whose conduct is under investigation is alleged to be a dominant player in the relevant market and is suspected of abusing its dominant position to the detriment of its competitors and the competition process.

The web search market is an example of a two-sided market where internet users account for one side and the advertisers for the other.⁵ Given the increased regulatory scrutiny faced by the web search market, this paper uses the two-sided market framework to analyze the market structure and the behavioral trends on both sides of the market in order to assess the state of competition in this market. Section 2 traces the evolution of the web search engines. Section 3 presents the two-sided market framework and examines trends on both sides of the web search market. Section 4 concludes.

II. The Evolution of Web Search Engines

The advent of the internet has made a vast amount of information available. From 26 known web sites in 1992, there are now over four million web sites and billions of web pages to browse.⁶ This rapid growth has given rise to the challenge of managing information so that users can find what they are looking for. Over the years many new products have been invented to help make the web easier to navigate, and one of the most useful of these products is the search engine.

Search engines are designed to search for information on the web by searching documents for user-specified keywords and returning a list where the keywords are found. Search engines generally consist of three main parts: a crawler pro-

gram that traverses the web and looks for webpages that are either not indexed or have been updated since they were last indexed; an index of sites that have been crawled; and a user interface that employs an algorithm to produce results to search queries.⁷ The evolution of web search engines since the 1990s highlights the efforts made to improve web crawling, indexing, and searching in order to make it easier to navigate the web.

In its infancy, the internet was simply a collection of sites that users could access to upload or download files; searching for a specific file meant navigating through each file. In 1990 Archie, one of the first attempts at organizing information on the Internet, was created. It provided a database of archived file-names, which it would try to match with users' queries. However, Archie did not index the content of the files. Another problem was that users had to manually

maintain the directory of sites that could be searched, which limited its reach.⁸

FROM 26 KNOWN WEB SITES IN
1992, THERE ARE NOW OVER FOUR
MILLION WEB SITES AND BILLIONS
OF WEB PAGES TO BROWSE.

The next defining step in web search engines was the introduction of robots, which automated the indexing system. The first search engine based on robot technology appeared in 1993. It

was called the World Wide Web Wanderer, and it collected information on websites and automatically added that information to an index. Although this robot technology significantly increased the number of sites that could be accessed through automation of the indexing, relevancy of the search results emerged as the next challenge.⁹

In 1994, spider technology was born. The older robots only indexed the sites and the titles of a page. In contrast, spiders (or crawlers) were software programs that indexed the entire content of a web page and recorded web links. The first crawler-based search engine, WebCrawler, appeared in 1994, which indexed not only the names and locations of websites but also their full text, making it possible to search within the text of web pages for desired information. This greatly improved the relevance of search results.¹⁰

Lycos, launched in 1994, was the first search engine to use hyperlinks between webpages to determine context and relevance. It displayed the title and ranking of a page, provided snippets of web pages, and added features such as prefix matching and word proximity. Yahoo! made its debut in 1994 as a directory, a list of categorized websites with search capability. Unlike other search engines, Yahoo! did not use spider technology to build automatic listings of websites. Instead, human editors were used to catalog the web. Consequently, its index of websites was quite small. In contrast, AltaVista, which was launched in 1995, was indexing up to ten million web pages a day. It did not rely on a single crawler program, and instead used thousands of crawlers to index the internet. It was the first high-speed search engine, the first to allow natural language queries and multi-lingual search, and it included features such as advanced searching tech-

niques (e.g. searching for phrases using quotes). It soon became a popular web search engine.¹¹

By 1997, several other search engines providing different degrees of innovation in web search had emerged. Excite, launched in 1995, used concept-based searching that utilized statistical word relationships, such as synonyms, to improve search results when the exact keyword was not entered. HotBot, released in 1996, made use of cookies to store personal search preferences, which enabled personalized search. Ask Jeeves, launched in 1997, used human editors to match users' search queries and ranked search results based on their popularity.¹²

IN 1998, GOOGLE
(INITIALLY CALLED
BACKRUB) WAS LAUNCHED.

In 1998, Google (initially called BackRub) was launched. It ranked web pages using citation notation, which monitored how many sites linked to a given web page. The more sites and the more important the sites that linked to a given web page, the higher was the site's ranking in the result list. Due to its unique high speed, combined with a unique relevancy based ranking of search results and a simple, easy-to-use interface, Google quickly became a popular search engine.¹³

Since 2000, several other search engines have appeared based on new search engine concepts and technology. For instance, Teoma, founded in 2000, had a unique link popularity algorithm that analyzed links, in context, to rank a web page's importance within its specific subject. For instance, a web page about "baseball" would rank higher if other web pages about "baseball" linked to it.¹⁴ In 2006 Snap was launched, with a completely transparent business model showing search volumes, revenues, and advertisers. It showed users how many others have searched for similar terms, and it also displayed search results with statistics like the number of user clicks and the average page views. In June 2009, Microsoft launched Bing, a new search service that changed the search landscape by providing a list of related searches directly in the result set.¹⁵

As the web continues to grow rapidly, the challenge of indexing the ever-growing web and producing relevant results to search queries has become enormous. Some search engines have emerged as all-purpose types and try to index the entire web; for instance, Google, Yahoo!, and Bing. Others have found their niche by narrowing their field to a specific field, language, or geographical region. For instance, Baidu, launched in 2000, is a popular Chinese web search engine and Guruji, launched in 2006, is India's first local search engine.¹⁶ Several job search engines, for example monster.com and job.com, have been established to allow employers to post job requirements as well as to help job seekers to search for suitable jobs.

Search engines seek to differentiate themselves on their comprehensiveness, up-to-datedness of their search index, and the relevance of their search results. Another means of differentiation is by providing features such as advanced

search, search in images, videos, maps, news and books, specialized search services such as Google Scholar for searching scholarly literature, text translation, and by providing additional services such as email.

Since the early days of web search, search engine providers have been active in two separate markets—not only in the search business, but also in the advertising business. In fact, advertisement is the main revenue source of many search engines including players such as Google and Yahoo!¹⁷

Advertising in the search engine context can take different forms. On the one hand, traditional types of advertisements—similar to those found in offline newspapers and magazines—such as display ads, sponsorships, and listings or classified ads have been replicated by search engine providers. On the other hand, search-specific advertising products have emerged. The two most prominent types are paid placement, where an advertisement is linked to a search term, and paid inclusion, where the advertiser pays a fee to the search engine provider in order to get a site included in the search index.¹⁸

Yahoo! was one of the first companies to monetize on-line search through advertising. It allowed advertisers to place banner ads on the search results page for a fee. But in the late 1990s, the search advertising industry was revolutionized by a purely commercial endeavor called GoTo (renamed Overture).¹⁹

GoTo solicited advertisements and indexed them by keyword. Users searched for relevant ads by keyword and the results returned were ranked based on how much the advertiser was willing to pay for the keyword. Advertisers bid on keywords, and advertisers only paid when a user clicked on the result.

SINCE THE EARLY DAYS OF WEB SEARCH, SEARCH ENGINE PROVIDERS HAVE BEEN ACTIVE IN TWO SEPARATE MARKETS—NOT ONLY IN THE SEARCH BUSINESS, BUT ALSO IN THE ADVERTISING BUSINESS.

This system introduced two new features to web advertising. First, the auction-based system allowed advertisers themselves to set a price on the keywords they valued. Second, advertisers only paid GoTo when a user clicked on the link associated with their ad. Before GoTo introduced the pay-per-click model, advertisers paid on cost-per-thousand-impressions model (“CPM”) by paying a set price for every thousand users who saw the ad. GoTo’s auctioned, pay-per-click method changed the way advertisers paid for online advertisements. Instead of paying for every thousand views of an ad that may or may not have been associated with a relevant search, advertisers paid only for actual clicks after a user searched for a specific keyword.²⁰

The GoTo advertising model has revolutionized online advertising for search engines. Unlike advertising on television or radio or on other forms of online media, search engines offer advertisers more than a general audience. Because

users of a search engine are actively looking for certain information, search engines are able to sell specific ads to advertisers. The GoTo model made search engines a commercially viable tool. By linking the search query term that the user types in with the advertisements that are displayed, search engines have found a way to help advertisers target precisely who they are looking for. This trend has given rise to two forms of search results: organic search results, which are generated through a search engine's own information sorting process; and paid search results, which are advertisements. These different forms of search results highlight the two-sided nature of a web search engine.

III. State of Competition

Web search is an example of a two-sided platform that enables two distinct but related groups of searchers and advertisers to obtain value that would not occur otherwise.²¹

A key feature of two-sided platforms is the presence of “indirect network effects.” As described by Evans, indirect network 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.²² A search platform is more valuable to advertisers if it has a large number of users. 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. Furthermore, advertising revenue enables a search platform to provide complementary products and services to users, such as email or photo sharing, which increases the value of the search engine for users.

A related feature of two-sided platforms is the need to “balance” the demands of the two sides. In setting prices, for example, 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. Thus, as Evans observed, for profit maximization any provider of a two-sided platform has to consider the demands of both sides, the interrelationships between these demands, and the costs of running the platform.²³ Furthermore, one side of a two-sided platform usually gets a better deal. For example, searchers do not pay search engines, but advertisers do. One of the reasons for this asymmetric pricing structure is that searchers give the search platform its value and create positive network effects.²⁴

ONE OF THE REASONS FOR THIS ASYMMETRIC PRICING STRUCTURE IS THAT SEARCHERS GIVE THE SEARCH PLATFORM ITS VALUE AND CREATE POSITIVE NETWORK EFFECTS.

Two-sided platform markets present unique practical challenges for antitrust analysis and enforcement. In a traditional market, the analysis centers on the

responses of a single set of customers to changes in, for example, price or output supplied and the responses of the suppliers to changes in demand. However, in a two-sided platform, market definition and market power analyses must take into account the possibility that the two sides of a platform are interdependent. Therefore, to assess the state of competition in web search engines, we consider both sides of the web search platform. Furthermore, web search engines are an example of a continuously evolving innovative market. Therefore, it would be useful to examine the state of competition in web search engines by examining the market structure and behavior of players over time. As Fisher observed, in antitrust analysis what matters are the constraints other firms and products put on the power of those whose actions are being examined.²⁵ Therefore, in our assessment we also consider the effect of the other web-based platforms, such as social networking sites on web search providers.

A. WORLDWIDE SEARCH MARKET

As per data published by comScore, more than eight hundred million people aged 15 years and over conducted worldwide web searches in January 2008, and the total number of worldwide searches was over one hundred billion in July 2009.²⁶

Figure 1 illustrates the pattern of search behavior across different parts of the world from August 2007 to July 2009. Among the five global regions, Europe accounted for the highest share of searches at 32 percent in July 2009, followed by Asia Pacific (31 percent) and North America (22 percent).

Figure 1

Worldwide Search by Region (share of searches)

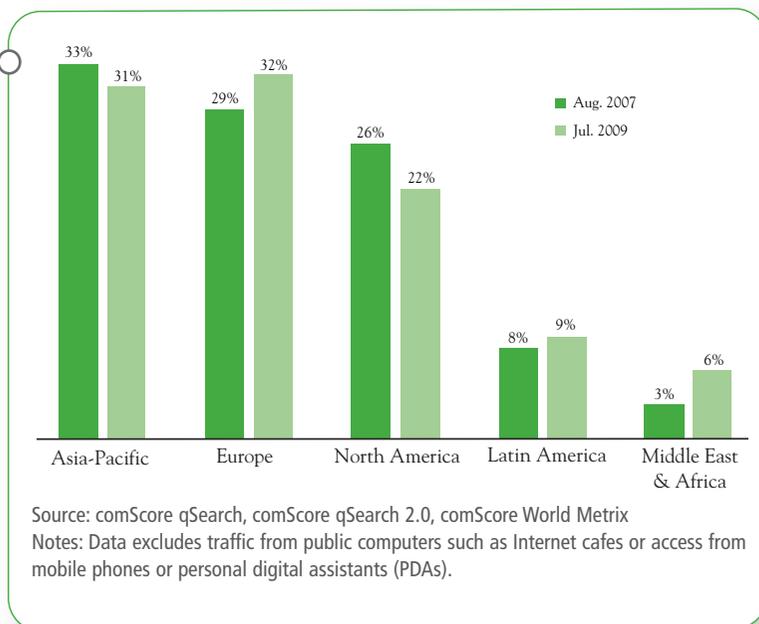


Figure 2 presents search behavior at the country level and shows that the United States remains the largest individual search market in the world with 22.7 billion searches, or approximately 17 percent of searches conducted globally in December 2009. China ranked second with 13.3 billion searches (ten percent share of worldwide search), followed by Japan with 9.2 billion (seven percent) and the United Kingdom with 6.2 billion (five percent). Germany, France, South Korea, Brazil, Canada, and Russia are the other top countries.

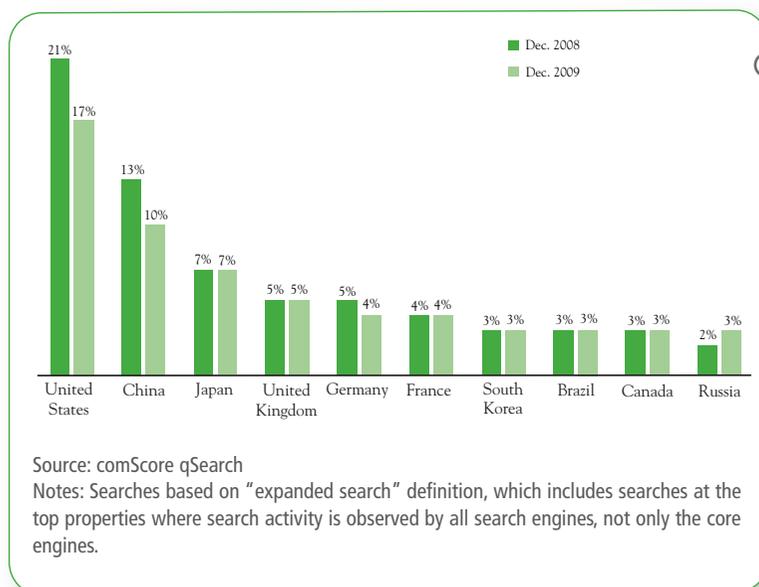


Figure 2

Top Ten Countries
by Share of
Searches

B. SEARCH SHARE BY SEARCH ENGINES

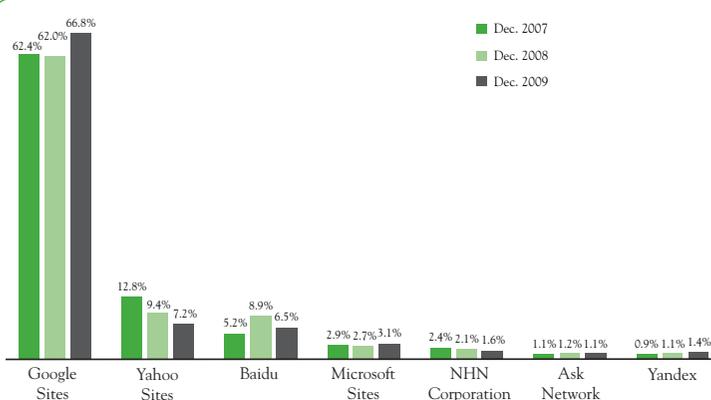
Figure 3 presents the data on top search engines by their share of searches worldwide. Google Sites ranked as the top search property worldwide with 87.8 billion searches in December 2009, or 66.8 percent of the global search market. Yahoo! Sites ranked second with 9.4 billion searches (7.2 percent of the searches worldwide), followed by the Chinese search engine Baidu with 8.5 billion searches (6.5 percent share). Microsoft Sites ranked fourth with four billion searches worldwide (3.1 percent), which increased from 2.4 billion searches in December 2008. The increase has primarily been attributed to the introduction of its new search engine, Bing, in June 2009. NHN Corporation, which owns Naver, the popular search engine in South Korea, ranked fifth with two billion searches (1.6 percent). Yandex's (the Russian search engine) share of global searches consistently increased during the three time periods, accounting for 1.9 billion searches worldwide in December 2009.

The data highlight that at the global level Google dwarfs the other search engines in searches. For the three time periods for which data are available,

Google's share in global searches increased while that of Yahoo!, the second ranked search engine, showed an opposite trend. The data also show the importance of country-level search engines, Baidu, NHN, and Yandex, suggesting that it is useful to discuss the search share at a country level.

Figure 3

Top Search Properties Worldwide by Share of Searches



Source: comScore qSearch, comScore qSearch 2.0

Notes: Excludes traffic from public computers such as Internet cafes or access from mobile phones or PDAs.

Table 1 reports search share data of leading search engines in the major countries in the four regions of Asia Pacific, Europe, North America, and Latin America for two different time periods, as well as the current top-ranked search engine.

The data show that Google is a dominant player in the North American region with approximately a seventy and eighty percent of search share respectively in the United States and Canada. It also occupies a dominant position in search queries in Brazil.

In Europe, Google is the leading search provider in the United Kingdom, Germany, and France with a search share of eighty to ninety percent in 2009/2010, which represents an increase in all the three countries as compared to the data recorded in 2007. Nevertheless, in Russia, Yandex is the leading search engine with a search share of 64 percent; this share has also increased from 2007.

Yandex's share of the Russian search market of 64 percent in December 2010 was far greater than Google's share of 22 percent.²⁷ There are several reasons for Yandex being the lead search engine in Russia. The focus on the Russian language helped Yandex occupy the lead position in the initial years as Google

struggled to adapt to the Russian language. However, Google has addressed this gap and has engineers based in Russia who fully understand the challenges of the Russian language. The key factors behind Yandex’s continued popularity appear to be its strategy of consistently adding and developing new technologies. For example, in late 2009, Yandex launched its “Matrixnet” machine-learning technology, which has significantly improved its search results by creating algorithms that “learn” as they are used and have increasingly complex ranking factors. Besides focusing on improving the quality of its search results, Yandex has been expanding into other related services. For example, in 2010, it acquired GIS technology to provide map services; it launched a job site in the same year; and it entered in a deal with Facebook.

Table 1

Leading Search Engines by Region and Country

Region/country	By share of searches				By rank ^a 2011 [^]
	2007 [@]		2009/2010		
<i>Asia Pacific</i>					
China	Baidu	54%	Baidu ^r	63%	Baidu
Japan	Yahoo!	49%	Google ^r	48%	Yahoo! Japan
South Korea	NHN (Naver)	65%	NHN (Naver) ^s	63%	NHN (Naver)
India	Google	81%	Google ^r	89%	Google
<i>Europe</i>					
U.K.	Google	74%	Google ^s	91%	Google
Germany	Google	80%	Google ^h	80%	Google
France	Google	82%	Google ^y	90%	Google
Russia	Yandex	52%	Yandex ^z	64%	Yandex
<i>North America</i>					
U.S.	Google	64%	Google ^u	72%	Google
Canada	Google	78%	Google ^r	80%	Google
<i>Latin America</i>					
Brazil	Google	90%	Google ^e	90%	Google

Source: comScore qSearch; comScore qSearch 2.0; comScore World Metrix; alexa.com (viewed 17 April 2011); LiveInternet (Russia); Hitwise (UK); Webhits (Germany); At Internet Institute (France); Koreanclick (South Korea); Evans (2008)²⁸.

Notes: Data excludes visits from public computers such as Internet cafes or access from mobile phones or PDAs.

^aThe sites in the top sites lists are ordered by their one month alexa traffic rank. The one month rank is calculated using a combination of average daily visitors and page views over the preceding month. The site with the highest combination of visitors and page views is ranked number one.

[@]Data are for December 2007; ^rData are for September 2009; [^]Data are for April 2011; ^sData are for May 2010; ^yData are for May 2008; ^hData are for June 2009; ^zData are for April 2009; ^uData are for December 2010; ^eData are for August 2010; ^zData are for July 2009.

In the Asia Pacific region, Google is the leading search engine in India with a search share of 89 percent as of September 2009. In Japan, Google took over from Yahoo! as the leading search engine in September 2009 only to be replaced

by Yahoo! Japan in April 2011. In the other two major Asian markets, China and South Korea, a domestic player has occupied the top spot in web searches.

Baidu is the most popular search engine in China with a 63 percent share of Chinese search. There are several reasons, commercial and regulatory, which contribute to Baidu's leading position. These include the censorship problems faced by Google in 2009, which until then had a 22 percent share of Chinese searches. Since then, Google's share dwindled to 11 percent in the fourth quarter of 2010, while Baidu's share increased to 84 percent. Baidu also occupies an edge over Google in returning precise search results of China domestic matters.²⁹

In South Korea, NHN Corporation owns Naver, the leading search engine. Its share of searches is over sixty percent; it is followed by Daum which has a share of 21 percent and Nate at third place with a share of nine percent.³⁰ Naver had been the dominant player in the South Korean search market for about a decade, but the market structure was different in the 1990s and in the early 2000s. Yahoo Korea, launched in 1997, occupied a dominant position until Daum was launched, enjoying a rapid growth on the basis of its Hanmail service which, by the beginning of 2000s, became the most popular search engine. However, this did not last long. Naver, launched in 1999, introduced Knowledge iN, a knowledge search service, together with integrated search and in 2001 it became one of the three leading search engines along with Daum and Yahoo by search query

volume. After 2003, Naver has been the most popular search engine.³¹ It is claimed that Naver's strength lies in its ability to understand the search culture of domestic users, which is reflected in its search results.³²

IT APPEARS THAT THE PRESENCE
OF GOOGLE AND YAHOO! AS
GLOBAL PLAYERS PERFORMING
IN A COUNTRY PUTS COMPETITIVE
PRESSURE ON THAT
COUNTRY'S DOMESTIC SEARCH
ENGINES TO PERFORM.

While Google is the leader in global searches, the market structures at individual country levels are quite different. In several countries, Google is the leading search provider, but in other countries domestic players occupy the top spot. It appears that the presence of Google and Yahoo! as global players performing in a country puts competitive pressure on that country's domestic search engines to perform. The leading domestic search engines have been responding to this competitive threat by making consistent efforts to develop new technologies, which is helping them maintain their lead. The rise of Google as the preferred global search engine can also be attributed to similar factors, which is demonstrated by the story of its ascendancy in the U.S. search engine market.

Figure 4 reports the shares of U.S. search traffic for the top five U.S. web search engines: Google, Yahoo!, MSN/Microsoft/Bing, Time Warner/AOL, and Ask. For the U.S. search market, in December 2004 Google and Yahoo! occupied comparable market shares of approximately 35 percent. Since then, however, Google has managed to increase its share to 64 percent while Yahoo! has seen

its market share plummet to 19 percent. However, Yahoo!'s search share has increased between 2009 and 2010, as has Microsoft's share since the launch of Bing in June 2009. Although Google's search share has declined between 2009 and 2010, it remains the leading search provider in the United States by a substantial margin.

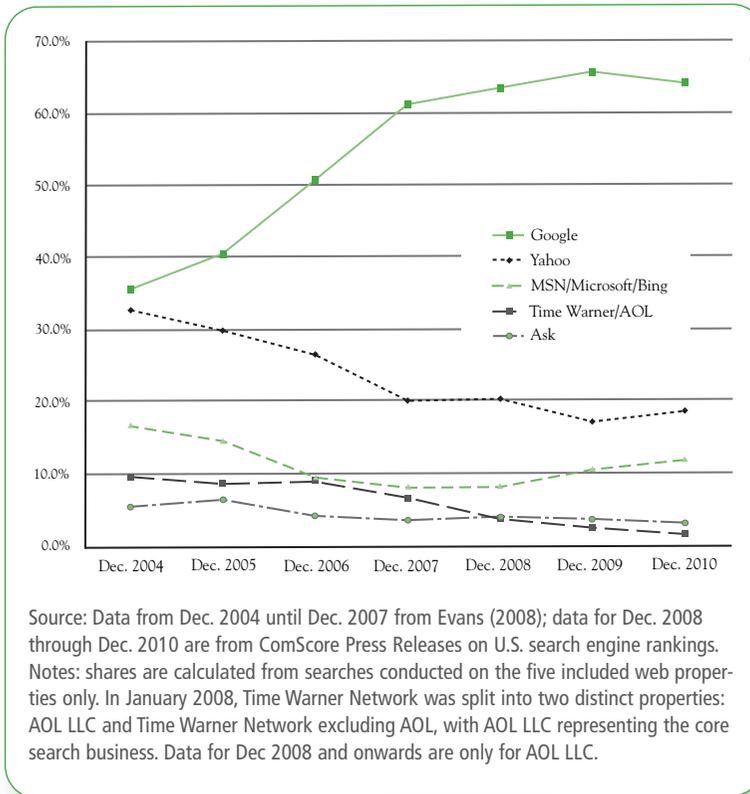


Figure 4

U.S. Search Engine Share by Searches

Table 2 reports the top three search providers in the United States from 1999 to 2007. It also indicates whether the firm used its own search engine technology or whether the technology was outsourced. Yahoo! held the top spot from 1999 to 2002; Google achieved a higher share than Yahoo! of search traffic in 2003 and has held the lead ever since. The table highlights that, Google, Yahoo! and MSN/Bing, the three global search engines, have emerged strongly from the churning in the web search market. By 2007, the search share of others such as AltaVista, Lycos, Ask Jeeves (now Ask), and Excite is miniscule.³³

Google's rise as a leading search provider demonstrates how a search engine can outperform its competitors based on superior innovation.³⁴ Its search algo-

Table 2

Leaders in Search in the United States by Search Traffic

Rank	1999	2000	2001	2002	2003	2004	2005	2006	2007
1 st	Yahoo! (Inktomi)	Yahoo! (Google)	Yahoo! (Google)	Yahoo! (Google)	Google (Own)	Google (Own)	Google (Own)	Google (Own)	Google (Own)
2 nd	AltaVista (Own)	AltaVista (Own)	Microsoft (Inktomi)	Google (Own)	Yahoo! (Google)	Yahoo! (Own)	Yahoo! (Own)	Yahoo! (Own)	Yahoo! (Own)
3 rd	Excite (Own)	Lycos (Fast Search & Transfer)	Google (Own)	Microsoft (Inktomi)	Microsoft (Inktomi)	Microsoft (Inktomi)	Microsoft (Own)	Microsoft (Own)	Microsoft (Own)

Source: Evans (2008)
Notes: Search technology powering search website given in parentheses; Yahoo acquired Inktomi in 2003.

rithm, which first incorporated web site popularity by taking into account the number of links pointing to a site, brought a significant increase in the quality of search results. This algorithm, called PageRank, is closely related to academic citation counting. It is based on the concept that the quality of an academic article depends on the number of other articles that cite it, and the quality of those citations depends on the number of citations generated by the citing articles. In web search, every link to a particular web site can be considered as being like an academic citation.³⁵ Thus, PageRank allowed Google to develop a way to index and search the internet that relied on a web page’s “reputation” with other pages rather than just on a page’s self-promotion. Its better organic search results drove users to it quickly in less than two years after it was launched, and it has maintained its lead by continuously striving to improve the quality of its search results. The recent gains in the search share of Yahoo! and Microsoft are due, in part, to the introduction of contextual search that tie content and related search results together. The various initiatives taken by the leading search providers are discussed in Section III(E).

C. TRENDS IN ONLINE ADVERTISING

Online advertising revenue has increased steadily over time, both in absolute terms and as a fraction of all advertising revenue. Evans (2008) reports that the share of U.S. online advertising expenditure to all advertising expenditure increased from 3.2 percent in 2000 to 8.8 percent in 2007. This trend has continued unabated as online ad spending in the United States at US \$25.8 billion in 2010 was estimated to surpass newspaper advertising, making it second only to TV advertising.³⁶ Furthermore, the online advertising market has proved resistant to the effects of the recession caused by the recent global financial crisis. Unlike spending on all other major media, worldwide online advertising spending increased from US \$54.2 billion in 2008 to US \$55.2 billion in 2009 and to US \$61.8 billion in 2010.³⁷

The relative mix of online advertising has also changed. Figure 5 shows the evolution of various online advertising formats from 2000 through 2010. In 2000, display advertising, which is similar to newspaper and magazine ads, accounted for 78 percent of total online ad spending, while search ads, which are linked to a search for a keyword, accounted for only one percent. Ten years later, in 2010, search ads accounted for the largest share of online ad revenue at 46 percent followed by display ads, which accounted for 38 percent. In the past decade, search-based advertising was the fastest-growing segment of online advertisement till 2004, and thereafter its growth has plateaued. This is mirrored by a decline in the share of display ads till 2006. Nevertheless, Figure 5 shows that since 2008 the share of display ads has consistently increased and the share of search ads has changed little. Together the two formats accounted for 84 percent of total U.S. online ad spending in 2010.

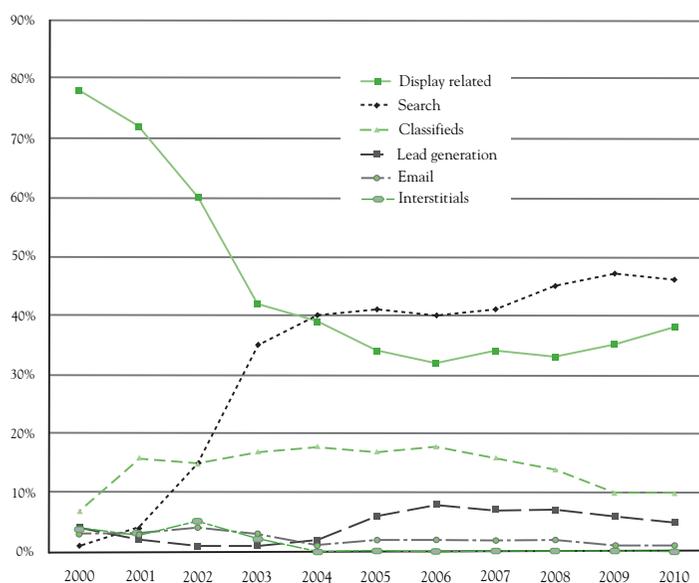


Figure 5

Share of U.S. Online Advertising Revenues by Different Formats

Source: Interactive Advertising Bureau press releases 2009-10; Evans (2008)

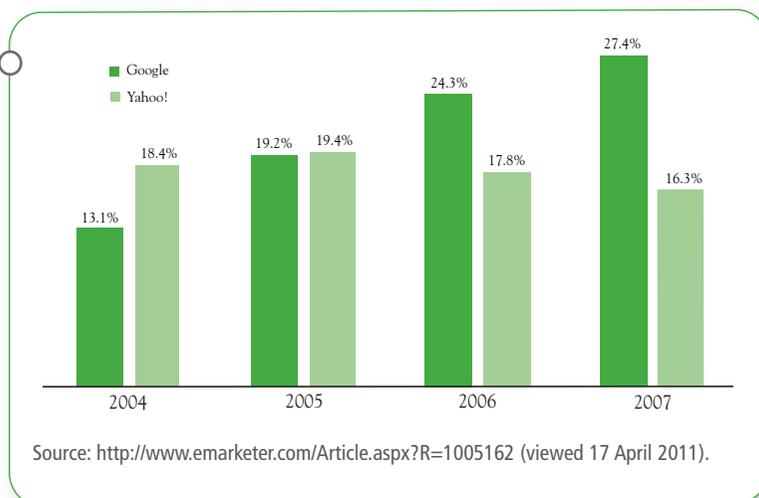
Notes: Display advertising refers to fees advertiser pays an internet company for space to display a static or hyper-linked banner or logo on one or more of the internet company's pages. Sponsorships represent custom content and/or experiences created for an advertiser that may or may not include ad elements such as display advertising, brand logos, advertorial, or pre-roll video. Email ads include banner ads, links, or advertiser sponsorships that appear in commercial e-mail communication. Interstitials are ads displayed during a transition from one web page to the next. Search refers to fees advertisers pay internet companies to list and/or link their company site domain name to a specific search word or phrase, and it includes paid search revenues. Classifieds refer to fees advertisers pay internet companies to list specific products or services (e.g., online job boards and employment listings, real estate listings, automotive listings, auction-based listings, yellow pages). Lead Generation refers to fees advertisers pay to internet advertising companies that refer qualified purchase inquiries (e.g., auto dealers that pay a fee in exchange for receiving a qualified purchase inquiry online).

D. SHARE OF SEARCH ENGINES IN ONLINE ADVERTISING

Figure 6 reports the U.S. online ad revenue shares of Google and Yahoo! from 2004 to 2007. Their combined share increased from 31.5 percent in 2004 to 43.7 percent in 2007. While Google's share increased steadily from 13.1 percent in 2004 to 27.4 percent in 2007, Yahoo!'s share declined to 16.3 percent in 2007 from a peak of 19.4 percent in 2005.

Figure 6

U.S. Online Ad Revenues at Google and Yahoo! (per cent of total U.S. online ad expenditure)



The data reported in Figure 6 are for the total online ad revenues. As discussed in Section III(C), display ads and search-based advertising are the two leading formats in online advertisement, and it will be useful to examine the ad revenue share of the search engines in these two formats. Data are, however, only available for display ads and Figure 7 reports the share of top U.S. internet publishers based on the number of display ad impressions delivered. Nearly half (48.8 percent) of all display ads seen by U.S. internet users originate on these properties.

In November 2007, Yahoo! sites ranked as the top display ad publisher property with 18.8 percent of display ad views, but by the third quarter of 2010, the popular social networking site Facebook led all online publishers with a share of 23.1 percent of online display ads, up from 1.5 per cent in November 2007. Yahoo! sites ranked second with a share of 11 percent, followed by Microsoft sites with 5 percent. Google's share of online display ads was 2.7 percent in the third quarter of 2010.

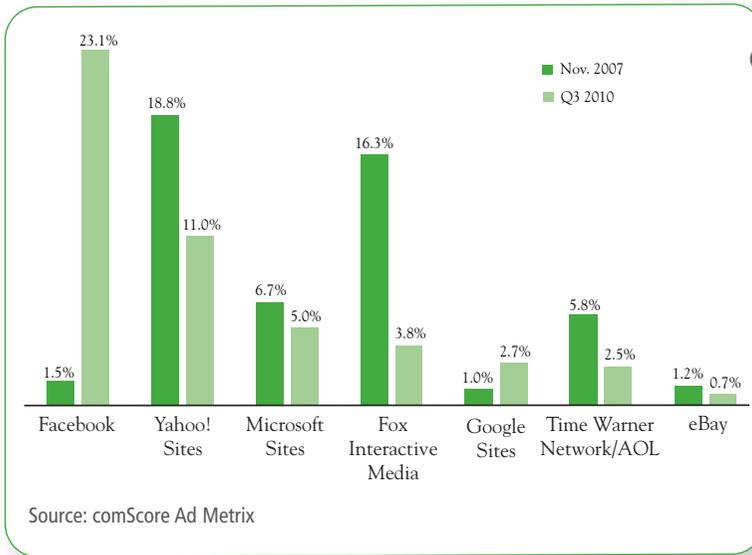


Figure 7

Top U.S. Online
Display Ad
Publishers by
Share of Online
Display Ads

Although the unit of measurement for data reported in Figures 6 and 7 are different—total online ad is reported as a revenue figure and the display ad format is reported as a number—Google’s relatively high share in total U.S. online ads and relatively small share in the display ad format seems to suggest that majority of its online advertisements are search-based. For advertisers, the greatest value comes from buying advertising on the search engine with the most users and the best results; that is, higher click-through rates. In both parameters Google stands out as a better platform. Google’s success with organic search and its dominance of the search market appears to have contributed to its success in getting a significant share of the search-based advertising market. The numbers thus seem to suggest Google’s linked dominance of both search and search-based online advertising.

However, Google’s dominance in online advertising appears to be under threat, and the most likely source for competition lies in the increasing popularity of social networking sites such as Facebook. It can be argued that since Google is already a small player in online display ads, the emergence of internet properties such as Facebook—a leading player in display ads—should not affect its bottom line. Nevertheless, as argued by Spulber,³⁸ advertisers allocate their expenditures on the basis of expected returns and web search providers compete with each other and with other types of media to attract advertisers. Besides social network websites, search providers are likely to face competition from other forms of emerging platforms, for example, tablets. Additionally, Goldfarb & Tucker³⁹ present empirical evidence of substitution between online and offline advertising. Thus, there appear to be indications of competitive threats on the advertising side of the web search engine market.

E. STRUCTURAL AND BEHAVIORAL RESPONSES

Search providers have adopted various strategies such as mergers and acquisitions (“M&A”) and the introduction of new technologies to improve their market share in search as well as online advertising. Table 3 provides a list of search providers ranked by their total M&A activity from 2000 to 2008. Yahoo!, Google, and SoftBank Corporation (which has a majority stake in Yahoo Japan) have been active in M&A.

Table 3

Mergers and Acquisitions by Search Providers (2000-2008)

Acquirer (including subsidiaries involved as acquirers)	Total number of mergers	Value of mergers (USD million)
Yahoo! Sites	53	6382.1
Google Sites	46	6477.0
SoftBank Corp (Yahoo Japan)	37	591.3
Overture Services Inc (Goto)	5	394.8
Microsoft/MSN	2	13.3

Source: Thomson Reuters' SDC Platinum

Approximately 76 percent of the acquisitions (by number) by the search providers were of target companies that provided internet-related services such as web search, online mapping technology and online document conversion. These acquisitions have helped the search providers expand their product line. For example, many of Google's well-known services are a result of acquisitions. This includes Google Docs (acquisition of Writely), Google Maps (acquisition of Keyhole), and Google's foray into mobile communications (acquisition of Android). Each of Google's acquisitions can be seen as a strategy to strengthen its market share either by attracting more advertisers, or by attracting more users to the search engine. Google's acquisition of DoubleClick is an example of an acquisition designed to expand its presence in the display advertisement market.

For several years Yahoo! outsourced its search service to other providers, considering it secondary to its directory and other content features, but by the end of 2002 it realized the importance and value of search and started aggressively acquiring search companies. Yahoo! acquired Inktomi in December, 2002, and Overture in July, 2003 (which had acquired AltaVista in 2003), and combined the technologies from these various search companies to make a new search engine. Consequently, as shown in Table 2, Yahoo! began using its own search technology in 2004.

Search providers have been focusing on introducing new features to their core business of search. For example, in June 2009 Microsoft launched Bing, a new search service that changed the search landscape by placing inline search suggestions for related searches directly in the result set. Yahoo! launched contextual search in 2005, which analyzes the page being read and gives a list of related search results. Therefore, instead of starting a search from a text box, a person would search while reading a specific page. Google launched an instant search interface in 2010 that suggests and displays search results while users type. Bing added Instant Answer to its image search results in 2010, which is a suggestion tool to help the users decide which definition of a query they want to see. Yahoo! introduced Search Direct in 2011 which is designed to provide users current relevant content, along with improved suggestions, and to display answers instantly to users as they type the search query in the search box.

IV. Conclusion

Web search is an example of a two-sided platform. In order to examine the state of competition in this market, it is important to consider the interdependence between its two sides, searchers and advertisers, and to identify the competitive constraints on both sides. It is also important to give due regard to the dynamic nature of competition in web search engines.

This paper shows that Google is the leading search provider globally and in many countries. There are, however, differences in market structures across countries. Furthermore, given the dynamic nature of the web search engine market, it is clear that a player's dominance depends on its innovation activity relative to others. Not surprisingly, search providers appear to be striving continuously to introduce new technologies to improve the quality of their search results. Online advertisements are the main revenue source for many search engines. With the advent of other forms of platforms that are likely to compete for online advertisements, it will be useful to examine their effects on search providers' strategies with respect to advertisements as well as search.

-
- 1 http://www.autoritedelaconurrence.fr/user/standard.php?id_rub=368&id_article=1514 (viewed April 6, 2011).
 - 2 L. Vitzilaiou, *Keywords: Google, European Commission. Anyone Feeling Lucky?*, 1(2) CPI ANTITRUST CHRON., 1-6, (January 2011).
 - 3 <http://www.reuters.com/article/2011/04/05/google-ftc-idUSN0420791620110405> (viewed April 6, 2011); R. Stutz, *An Examination of the Antitrust Issues Posed by Google's Acquisition of ITA*, American Antitrust Institute White Paper (February 2011), <http://www.antitrustinstitute.org/sites/default/files/Google-ITA%20AAI%20White%20Paper2.18.11.pdf>; *Google Antitrust Probe is Urged by Republican U.S. Senator* (<http://www.bloomberg.com/news/2011-03-11/google-antitrust-investigation-urged-by-republican-u-s-senator.html>, viewed April 6, 2011)

- 4 S. Ning, A. Ng, & L. Jia, *China: Wiki-Hudong against Wiki-Baidu - an abuse of dominance?*, <http://www.mondaq.com/article.asp?articleid=126928> (viewed April 6, 2011).
- 5 Vitzilaiou, *supra* note 2.
- 6 P. Iyer & B. Courtney, *Google: Online Search & the Battle for Clicks*, Sloan Fellows Program in Innovation & Global Leadership (May 2005); <http://www.dmoz.org> (The Open Directory Project, viewed April 9, 2011).
- 7 K.L. Devine, *Preserving Competition in Multi-Sided Innovative Markets: How Do You Solve a Problem like Google?*, 10(1) N. CAROLINA J. LAW & TECH., 59-118 (2009).
- 8 Iyer & Courtney, *supra* note 6.
- 9 *Id.*
- 10 Devine, *supra* note 7.
- 11 U. Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, (9) YALE J. LAW & TECH., 201-234 (2006).
- 12 *Id.*
- 13 Devine, *supra* note 7.
- 14 <http://en.wikipedia.org/wiki/Teoma> (viewed April 15, 2011).
- 15 Iyer & Courtney, *supra* note 6.
- 16 <http://www.guruji.com/Press/India's%20first%20local%20internet%20search%20engine%20launched.html> (viewed 10 April 2011).
- 17 Gasser, *supra* note 11.
- 18 *Id.*
- 19 Devine, *supra* note 7.
- 20 *Id.*
- 21 Vitzilaiou, *supra* note 2.
- 22 David S. Evans, *Two-Sided Market Definition*, ABA Section of Antitrust Law, Market Definition in Antitrust: Theory and Case Studies (2009), available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1396751.
- 23 *Id.*
- 24 *Id.*, Vitzilaiou, *supra* note 2.
- 25 F. M. Fisher, *Horizontal Mergers: Triage and Treatment*, 1(2) J. ECON. PERSPECTIVES, 23-40 (1987).

- 26 http://www.comscore.com/Press_Events/Press_Releases/2009/8/Global_Search_Market_Draws_More_than_100_Billion_Searches_per_Month (viewed April 10, 2011).
- 27 <http://blog.quintura.com/2010/04/12/yandex-boosts-russian-search-market-share-to-63-could-resume-ipo-plans/>; <http://searchengineland.com/yandex-keeps-on-beating-google-in-russia-62186> (viewed April 18, 2011).
- 28 David S. Evans, *Antitrust Issues Raised by the Emerging Global Internet Economy*, 102(4) NORTHWESTERN L. REV., 1987-2007 (2008).
- 29 <http://www.resonancechina.com/2010/08/23/introductions-to-china%e2%80%99s-popular-search-engines/>; <http://www.resonancechina.com/2011/03/14/baidu-83-6-search-query-market-share/> (viewed April 18, 2011).
- 30 <http://searchengineland.com/should-korean-search-engine-naver-worry-about-local-competitors-or-google-65401> (viewed April 18, 2011).
- 31 <http://www.koreaitimes.com/story/3083/naver-myth-internet-age-success> (viewed April 18, 2011).
- 32 <http://www.koreaitimes.com/story/3083/naver-myth-internet-age-success> (viewed April 18, 2011).
- 33 <http://www.searchenginehistory.com>
- 34 Devine, *supra* note 7.
- 35 D.F. Spulber, *The Map of Commerce: Internet Search, Competition, and the Circular Flow of Information*, (5) J. COMPETITION L. & ECON., 1-37 (2009).
- 36 http://www.emarketer.com/Report.aspx?code=emarketer_2000725 (viewed April 6, 2011).
- 37 http://www.emarketer.com/Report.aspx?code=emarketer_2000710 (viewed April 6, 2011).
- 38 Spulber, *supra* note 35.
- 39 A. Goldfarb & C. Tucker, *Substitution between offline and online advertising markets*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1721001 (November 2010).



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Payment Card Regulation and the Use of Economic Analysis in Antitrust

Jean Tirole
Univ. of Toulouse

Payment Card Regulation and the Use of Economic Analysis in Antitrust

By Jean Tirole*

A key input of our modern economies, payment cards are ubiquitous; debit and credit cards offer a wide range of alternatives to cash and checks to operate brick and mortar, e- and mobile phone, and P2P payments. The contours of the industry are rapidly changing.

The payment card industry is also becoming one of the most heavily regulated industries in some parts of the world. The United States and Europe, as well as a number of other jurisdictions across the world, have been or are in the process of regulating, *inter alia*, the network-determined payment made by the merchant's bank (called the acquirer) to the cardholder's bank (the issuer). This "Interchange Fee" has been the object of much controversy and the theoretical underpinnings of its regulation are still debated. The primary object of this note is to clarify the considerations that should be brought to bear on the determination of regulated fees. It argues that some broadly contemplated regulatory methodologies bear only limited resemblance with economically sound precepts. Finally, it derives some implications of these regulations for the likely evolution of the payment card industry.

*Jean Tirole is chairman of the board of the Toulouse School of Economics and scientific director of the Institut d'Economie Industrielle, University of Toulouse 1 Capitole. He is also affiliated with MIT, where he holds a visiting position.

I. Introduction

The card payment industry is remarkable in several respects and definitely worthy of our attention. First, it is sizeable.¹

Second, its contours are rapidly changing. The competition between “four-party,” mainly open systems such as Visa and MasterCard, and “three-party,” mainly closed² systems such as American Express and Discover rages, with old-fashioned payment means (cash, checks) still commanding a respectable market share.³ It is hard to predict who, among incumbents or entrants, will end up serving the various and interrelated business segments: debit and credit transactions; large, mid-size, and micro payments; e- and mobile phone payments; P2P payments; and so forth.

Third, this experimentation with alternative business models and the subsequent shakeout will be deeply affected by regulation, as the payment industry is becoming one of the most intensively regulated industries in some parts of the world.

Fourth, and of particular interest to antitrust practitioners and economists, the industry’s two-sidedness and other specificities make received antitrust doctrine largely irrelevant, and fresh thinking is required in order to design sound policy intervention.

The purpose of this note is to guide the reader through the intricacies of a rigorous understanding of regulatory stakes and interventions, and to hopefully clarify an otherwise muddled debate.

II. Description

A. THREE- AND FOUR-PARTY SYSTEMS

We first describe the two dominant business models of the card payment industry.⁴

The basics of four-party payment systems such as Visa and MasterCard can be grasped by looking at Figure 1. A card transaction between a merchant and a consumer is enabled by two system member banks, the acquirer (the merchant’s bank) and the issuer (the cardholder’s bank). For each card transaction the issuer and the acquirer pay system fees to the network. Because these fees are not at the core of current disputes we will ignore them in the following.

When a consumer makes a debit card purchase of 100 at a merchant, the merchant transfers this information to her bank, the acquirer, who credits the merchant say 99 if the merchant fee (or “discount”) is 1 percent. The acquirer uses this discount to cover his acquiring cost, his margin, the fee paid to the system (Visa or MasterCard), and finally—and also the focus of the regulatory attention—the interchange fee (“IF,” often called in the European context “MIF” for “multi-lateral interchange fee”) paid not to the system but to the cardholder’s

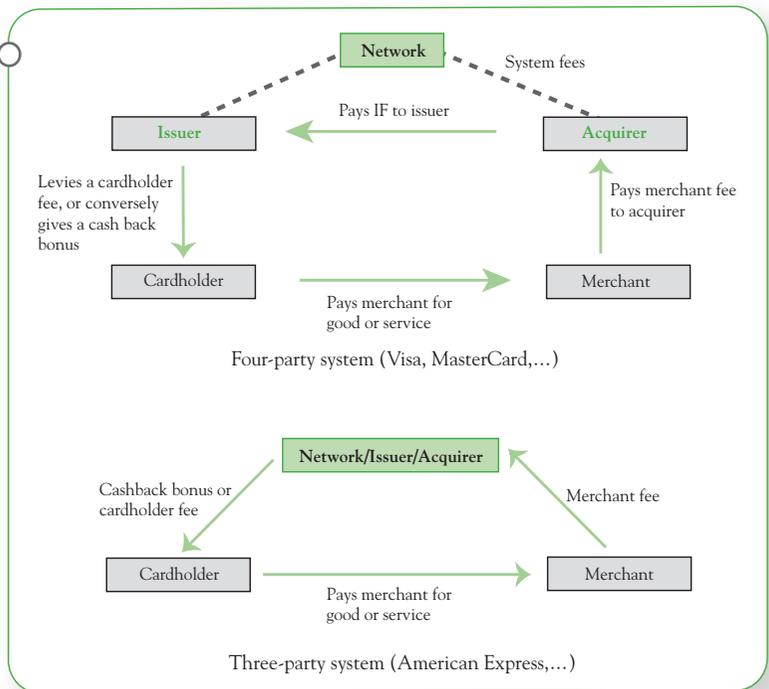
bank, the issuer. The issuer in turn debits 100 from the cardholder's account and may either charge the cardholder for the transaction or reward him through frequent flyer miles, cash back bonuses, or another reward instrument.⁵ The merchant fee is equal to the IF plus the acquiring cost (plus the system fee) if the acquiring industry is competitive.

The IF level may vary within a country by merchant category, merchant size, type of payment instrument (Visa/MasterCard, PIN debit/signature debit/credit, premium/basic, etc.), and then varies across countries. For example, large supermarkets were brought on board through a reduction (roughly by half in the United States) of IFs applicable to them. And, over the last decade, quick service restaurants such as large hamburger and pizza chains, which, in view of their \$5 or \$6 bills, relied exclusively on cash payments, were induced to take the Visa and MasterCard through tailored merchant pricing of card payments. Card payments now account for close to half of quick service restaurants' sales in the United States.

Similarly, IFs and merchant fees do vary within a card system depending on the reward system. High-reward (premium) cards can command an IF double that of basic cards.⁶ Merchants are prevented by the honor-all-cards rule (and also perhaps by transaction costs) from picking a subset of cards within a system.

Figure 1

Four- and Three-Party Systems



In (pure) three-party payment systems, the network, acquirer, and issuer are a single entity. So there is no explicit IF. One can define an implicit or *shadow IF*, though, by considering a fictitious competitive acquiring industry within the system. Because American Express can delegate its acquiring services to a competitive industry and achieve exactly the integrated outcome by setting an interchange fee, this shadow IF is equal to the difference between American Express' merchant fee and its acquiring cost.

IN (PURE) THREE-PARTY PAYMENT SYSTEMS, THE NETWORK, ACQUIRER, AND ISSUER ARE A SINGLE ENTITY. ONE CAN DEFINE AN IMPLICIT OR SHADOW IF, THOUGH.

Regardless of the organizational form, the merchant may or may not surcharge for card usage. Most often, the merchant does not surcharge and the customer pays the same amount to the merchant regardless of the payment method. This absence of surcharging may be due to a combination of wanting to induce card payments or to attract customers inclined to pay by card, of transaction costs associated with surcharging, and—in some countries and card systems—of an explicit prohibition to surcharge (although these prohibitions are gradually disappearing due to regulatory intervention. Australia and the United States, following the RBA investigation and the Visa/MasterCard settlement with DOJ, respectively, are cases in point). As Prager et al.⁷ point out, the fact that only a small number of merchants surcharge *per se* does not imply that surcharging is broadly irrelevant; it might be the case that the threat of surcharging effectively caps the merchant fee. Nonetheless, past experience shows that even high merchant fees need not lead to surcharging.⁸

B. IF REGULATIONS

These have been many investigations into IF setting since the late 70's and the *NaBanco* case against Visa⁹ in the United States. The last decade has witnessed particularly intense antitrust and regulatory questioning of the IF level in most developed countries, with Australia as an emblematic case of mandated¹⁰ reduction of IFs for Visa and MasterCard.¹¹

IF regulation has sometimes been motivated by the associated agreement among competitors (the issuers). This “illegal-price-fixing” argument, which was the basis for the *NaBanco* case and was invalidated by the courts in 1984, is based on an incorrect analogy. An increase in the IF is not a price increase for some final users like in standard cartel theory, but a reallocation of cost between two categories of end-users (merchants and cardholders). This point was made by authorities' staff in some regulatory hearings, and yet is not always taken on board as a key principle for policy intervention.¹²

United States: Following the lead of the Reserve Bank of Australia,¹³ which uses an issuer-cost-based approach to compute a cap on the IF, the recent regulatory proposals in the United States follow a “cost of service” or “public utility” approach. The Dodd-Frank Act requires the Federal Reserve Board to issue reg-

ulations on debit-card interchange fees and stated that they should be “reasonable and proportional to cost incurred by the issuer with respect to the transaction.” According to some estimates¹⁴ the proposed regulations will wipe out between 75 and 85 percent of issuers’ debit card interchange fee revenue.

Europe: The European Commission recently fixed an IF for MasterCard cross-border transactions equal to 0.2 percent on average for debit card transactions and to 0.3 percent for credit card transactions.¹⁵ Cross-border transactions represent a small fraction of total card transactions in Europe, but this cross-border rule is widely expected to impact the IF levels for domestic markets. In contrast with American regulators, the European Commission has chosen to regulate the cross-border IFs in such a way that the merchant fee does not exceed the retailer’s avoided-cost when a cash (or check) payment is replaced by a card payment:¹⁶

“But most importantly the aim of the new MIFs to be applied by MasterCard is that a merchant’s costs in accepting card payments should be no higher than the benefits from avoiding receiving cash.”¹⁷

This “cash-substitution” approach is closely related to the “tourist test,” which we later discuss. The European Commission further requires the methodology to be “transparent.”¹⁸

In the following, we will discuss whether these methodologies are sound, not the particular numbers that agencies or parties may come up with in application of the regulatory methodology.

III. Looking for a Market Failure

Basic economics by no means vindicates *laissez-faire*—market failures abound—and that fact offers substantial scope for improvement through sensible policy intervention. But basic economics also teaches us that policy interventions must be grounded in a rigorous treatment of several questions: What is the exact market failure and is it sizeable? Does the state have the information and the instruments to correct the failure? Will the remedy’s costs be offset by sufficient benefits? Such questions should be satisfactorily investigated before enacting new regulations.

The payment industry is no exception to the rule. Because there is widespread confusion about where the market failure lies,¹⁹ we start by identifying it. It is sometimes believed that the joint determination of an IF by banks represents an attempt to cartelize and raise prices. Economists and antitrust enforcers are rightly suspicious of attempts by competitors to get together and raise prices to users. The snag with this reasoning in the case of payment cards, though, is that there

are two groups of users and that increasing the IF raises the price of card transactions for one group (merchants) and lowers it for another (cardholders). Put differently, in a first approximation²⁰ the IF affects the price *structure* and not the price *level*.²¹ This feature by itself makes received knowledge about “cartelization” inadequate.

Later we will remark on the implications of banning the collective setting of IFs. For most of our analysis, let us follow regulatory practice, keep the institution as a given, and examine the consequences of *laissez-faire*. We will also assume that merchants’ surcharging for card usage is either prohibited or deemed by merchants to involve high transaction costs or to be unattractive to the consumer; when retail prices are the same regardless of the means of payment, the price of a card transaction relative to a cash transaction is zero. These two assumptions are broadly realistic in our current environment.

PUT DIFFERENTLY, IN A FIRST APPROXIMATION THE IF AFFECTS THE PRICE *STRUCTURE* AND NOT THE PRICE *LEVEL*.

There accordingly can exist two externalities between end-users:

- The merchant may not take a card that the cardholder would like to use, implying a loss of surplus by the latter.
- Conversely, the cardholder may prefer a means of payment that the merchant finds costlier than an alternative payment method. The cardholder then does not internalize the extra cost he imposes on the merchant.

To be sure, there are limits to such externalities: The merchant’s refusal to accept a card makes her business less attractive to consumers, implying a loss in goodwill; conversely, if cardholders want to use cards that are very costly to the merchant, the latter always has the option of rejecting it.²² Still, in a world in which the choice of payment method is unpriced,²³ nothing guarantees that the end-users’ decisions (acceptance, usage) are “socially right” in that they maximize joint user surplus.

Interestingly, the magnitudes of these externalities are determined by the IF. In the relevant range of IFs, one externality increases while the other decreases when the IF moves around. For example, suppose that the IF and, consequently, the merchant fee are sufficiently high that a card payment is more expensive for the merchant than a cash payment (the reader may wonder why the merchant then keeps accepting the card—more on this shortly). An increase in the IF raises the merchant fee, makes cards less attractive to merchants, and reduces the fraction of shops that take the card, depriving cardholders of the ability to use their preferred payment method. Concurrently, those merchants who keep accepting the card are hurt more badly when cardholders use the card. One thus sees the prominent role of IFs and understands why they command so much attention from merchants and policymakers.

Finally, and in order to streamline the presentation, we will abstract from issuer and acquirer market power and later explain how relaxing this assumption affects the analysis. It is generally considered that the acquiring industry, despite some concentration, is, in a number of countries, rather competitive.²⁴ Acquiring services are pretty close substitutes; because merchants are relatively well informed and eager to shop around for the lower fee, this segment of the industry in countries with competitive acquiring is rather commoditized. By contrast, the issuing side, despite much entry and numerous competitors, may be less competitive in the short run.

IV. Merchants' Demand for Card Payments

A. TWO BENCHMARKS

There is much confusion about the measurement of how much merchants are "willing to pay for a card payment," that is about their demand for card payments. Dispelling this confusion requires defining the alternative to a card payment: Would the payment be made through "cash" (broadly construed to include cash and checks, the traditional means of payment) instead? Or would there be no transaction and no payment at all?

Even focusing on the first alternative, cash substitution, measuring the merchants' willingness to pay also requires making a distinction between:

- The net benefit that merchants enjoy when the customer uses a card rather than cash, and
- How much they are willing to pay once they *further* take into consideration their desire to attract the customer.

The two notions coincide when the consumer enters the shop without knowledge of its card acceptance policy, stands captive at the cash register (is confronted with the set of payment options dictated by the merchant), and is able to pay by cash (or check). Hence the terminology of the "tourist test" coined in my paper with Jean-Charles Rochet,²⁵ a test which was much discussed in the regulatory hearings at the European Commission and at the Federal Reserve Board: "Does the IF level lead to a merchant fee that would induce the merchant to turn down the card for a tourist who has cash, assuming that the merchant had this discretion?"

DOES THE IF LEVEL LEAD
TO A MERCHANT FEE THAT
WOULD INDUCE THE MERCHANT TO
TURN DOWN THE CARD FOR A
TOURIST WHO HAS CASH,
ASSUMING THAT THE MERCHANT
HAD THIS DISCRETION?

In general, though, the second notion exceeds the first. Consumers may inquire into whether the shop takes the card before going to or entering the shop; or else consumers may be repeat consumers. Either way, accepting the card

makes the shop more attractive and results in extra sales. The merchant may then take the card even though she wished that, conditional on the customer making a purchase, that purchase be made through an alternative means of payment such as cash or check. Put differently, the merchant is willing to pay the high level *ex ante* (before the customer decides whether to visit the store) but not *ex post* (once the customer is captive). It is clear for example that merchants' acceptance of American Express cards at fees of 3 or 4 percent of the transaction (as was the case for a long time) was motivated in part by the desire to attract the (generally well-to-do) customers who carried American Express cards rather than by just a demand for economizing on the cost of cash transactions, however high this cost may be. Table 1 lists various elements that enter the narrow and broad concepts.

Let S denote the merchant's net benefit/cost savings from card usage, and A the attractiveness benefit. Let c_a denote the cost of the transaction for the acquirer. An IF equal to $S - c_a$ leads to a merchant fee equal to S once competitive acquirers have added their acquiring cost to the IF. The merchant is then *ex post* indifferent between a cash payment and a card payment. By contrast, in order to attract a customer, she will accept merchant fees up to $S+A$, that is take cards for which the IF does not exceed $S+A - c_a$.

Narrow concept: S	Broad concept: $S+A$	Table 1
Merchant's net benefit from card usage (tourist test)	Inclusive of attractiveness benefit	Merchant's Willingness to Pay
1. <i>Cash substitution</i> : card usage <ul style="list-style-type: none"> • Eliminates handling/depositing cost • (If network offers payment guarantee eliminates cost of fraud (counterfeit money, invalid checks)) • Reduces risk of holdup in store and employee theft, and economizes cost of armored vehicle transfers (when cash is substitute) • Reduces delay at counter (mainly when check is substitute) • Reduces float/speeds up crediting of merchant's bank account (relative to a check) • Provides merchant with value-added services; e.g., facilitates tracking of consumer purchases/reporting services 2. <i>Missed sales</i>	Narrow concept, <i>plus</i> <ol style="list-style-type: none"> 1. <i>Cash substitution</i>: card provides cardholder with <ul style="list-style-type: none"> • Convenience benefit of card usage for consumer (no need to go to ATM, or in case of a large amount to bank branch) • Interest-free period for consumer (differed debit cards) • Flexibility in managing cash balance (credit cards) • Payment traceability 2. <i>Missed sales</i> 	

The first entry in the narrow and broad concepts is usually referred to as “cash substitution,” although “cash” should obviously be taken to mean “cash and

checks;” its ingredients are straightforward. By contrast, the second “missed sales” entry in the narrow and broad concepts requires some explanations.

Suppose, first, that a customer in the shop does not have enough money in his bank account to purchase the good or service immediately. Either the purchase was unforeseen or the transaction costs of asking for an overdraft facility at his bank were perceived as high. Were the merchant not to accept credit cards, the transaction would not take place, generating a loss for the cardholder, but also for the merchant, who would then lose the (usually substantial) markup on the retail good.²⁶ The avoidance of missed sales is therefore an important benefit for the merchant. There is indirect evidence that this benefit may be sizeable: For durable goods, large merchants often subsidize credit on their own despite the likely inefficiencies involved (such as additional transaction costs stemming from a new loan contract and the multiplication of the consumer’s creditors, making the surveillance of consumer solvency more difficult).

Even for debit cards, there may not be any possibility of payment by check or cash, resulting in missed sales. For a brick and mortar outlet, the (time-pressed) customer may have no checkbook with him and no ATM easily accessible; or the shop (flower shop for instance²⁷) may be reached by phone. More importantly still, e-commerce is vastly facilitated by the use of electronic payments. Cash or even checks really cannot easily substitute for cards for online purchases.

Either way, the substitute for a card transaction in such situations is no transaction at all rather than a cash transaction.

The possibility of missed sales, which, in practice, seems quite important for the parties, has unfortunately been underexplored in the economics literature, which focuses on the cardholders’ choice of merchant and on the joint determination by cardholders and merchants of the payment instrument through which

THE POSSIBILITY OF MISSED SALES,
WHICH, IN PRACTICE, SEEMS QUITE
IMPORTANT FOR THE PARTIES,
HAS UNFORTUNATELY BEEN
UNDEREXPLORED IN THE
ECONOMICS LITERATURE.

they will transact.²⁸ Thus missed sales are an important topic of reflection in the agenda of decision-makers and academics (theory, empirics), but we can make a few tentative points nonetheless.

The socially relevant question is whether end-users (cardholders, merchants) exert externalities on each other and, if so, how we can induce them to internalize these externalities. Put differently and anticipating somewhat, one should ask (a) what those externalities are, and (b) how the IF can be used to make the parties internalize the externalities they impose on each other.

A higher IF on credit cards, for example, leads to cheaper credit cards for the consumers and encourages them to hold and carry such a card. This creates a benefit for the merchant, which enters the narrow definition. The possibility of

missed sales also enters the extra, attractiveness term of the broad definition; for example, suppose that a consumer holds a credit card and, having insufficient liquidity on his bank account, contemplates a credit purchase at the merchant. Such a purchase is infeasible—or costly to implement for the consumer—if the merchant turns down the credit card. So credit card acceptance contributes to the merchant's attractiveness.

Our treatment below will be couched in terms of cash substitution but, as noted above, missed sales matter.

B. NETWORK CHOICE OF IF

Next, we can inquire into what IF a payment card network would like to set. Suppose that the network aims at maximizing card volume.²⁹ Because a higher IF makes card usage more attractive to cardholders, the network should set the highest possible IF, i.e. the highest IF consistent with the merchants not rejecting the card. This is where our two benchmarks come into play. First, the IF can, in no circumstance, exceed $S + A - c_a$, as this would always lead merchants to turn down the card. Conversely, any IF below or equal to the tourist test level $S - c_a$ always leads to merchant acceptance, since card transactions then minimize the cost of transactions for the merchant and furthermore card acceptance may attract new customers.

Where in the interval $[S - c_a, S + A - c_a]$ will the network set its IF? The answer to this question hinges on two factors:

- *Consumer information about card acceptance:* Suppose all consumers are “tourists” in the sense that they are unaware of card acceptance policies when deciding where to shop. Then the network cannot charge an IF above the tourist test (or cash substitution) level $S - c_a$. By contrast, with well-informed consumers, the IF can be set at $S + A - c_a$ without inducing merchants to reject the card.
- *Cardholders' number of cards/systems:* Suppose that consumers have, say, two cards in their pocket,³⁰ one issued by a member of the Visa network and one issued by a MasterCard member. They are then said to “multi-home.” If the Visa network charges a higher IF than the MasterCard network, the merchant finds Visa cards more expensive than MasterCard's and stop accepting Visa cards whenever the Visa merchant fee exceeds S .³¹ Suppose that attractiveness is a concern for the merchant, and so the merchant internalizes the customer's net benefit from card transactions;³² the merchant then aims at inducing a choice of payment method that doesn't maximize her own direct benefit, but rather the joint surplus of the merchant and the customer.³³ The closer the merchant fee is to the merchant's net benefit S , the smaller the externality of the customer's choice of payment method on the merchant, and thus the better the customer's decision from the point of view of the joint surplus.

The importance of multiple-card-holding is highlighted by the downward pressure on American Express' merchant fee upon the introduction of no-fee-cards in the United States in the early 90's.³⁴ Numerous MasterCard and Visa cards were offered that involved no yearly fee. American Express cardholders could costlessly use such cards as a backup for their Amex card and so merchants were less reluctant to turn down the Amex card, which carried very high merchant fees.³⁵

NOTE, FINALLY, THAT SYSTEM
COMPETITION BRINGS IFs DOWN
ONLY IF CARDHOLDERS MULTI-HOME
ON MULTIPLE SYSTEMS.

Note, finally, that system competition brings IFs down only if cardholders multi-home on multiple systems. If cardholders hold a single card or hold several cards on the same system (e.g. several Visa cards), system competition exerts no pressure on the IF, as only the cardholder has a choice. The merchant has no choice but to accept the cardholder's system offer if she wants to transact by card with the cardholder. This situation is known as the "competitive bottleneck" case in the economics literature.

To sum up, consumer information about the merchants' card acceptance policies drives the IF up and away from its tourist test level. By contrast, the consumers' holding cards from multiple systems drives the IF down towards its tourist test level.

V. What Does Economics Say About IF Regulation and Is Current Regulation Economically Sound?

A. THE TOURIST TEST AS A BENCHMARK FOR REGULATION

In an industry fraught with externalities there is no guarantee that private decisions achieve a socially satisfactory outcome. Indeed, in the simple world described so far, it is easily seen that private interests can only lead to an IF, and consequently a merchant discount, that are higher than what society would desire. A basic economic precept is that welfare optimization requires economic agents not to exert externalities on each other. Suppose that the IF is set at its tourist test level. Then the merchant, by definition, is indifferent as to the choice of payment method by the consumer; there is, therefore, no externality and the consumer makes the socially correct decision. Thus in the simple world considered so far and in the absence of other distortion (such as issuer or acquirer market power, or improperly priced alternative means of payment) the socially optimal IF is equal to its tourist test level.

This reasoning assumes that merchants have the same cost savings from card payments S . When merchants differ in their cost savings, the proper generalization of this rule (still assuming that there is no market power) is that the card-

holder internalizes the average cost savings among merchants who take the card.³⁶ This rule implies that at the social optimum those merchants who receive the lowest cost savings (less than average) among those who take the card will fail the tourist test; for them accepting the card increases cost. To take an example, suppose that there are three categories of merchants, with card cost savings equal to 1, 2, and 3 percent of the transaction, respectively, and there are equal numbers of each category. The socially optimal IF is then 2 percent. This requires that merchants attribute a value of at least 1 percent to attractiveness, so that the category with only 1 percent cost savings is kept on board when the IF is 2 percent.

Can the system-optimal IF exceed its socially optimal level? The answer is a clear “yes.” As we just saw, the merchant may reluctantly take a cost-increasing card so as to attract customers. This attractiveness concern may, in theory, result in too many card transactions,³⁷ justifying John Vickers³⁸ use of the “must-take card” terminology.

Section VII will qualify this analysis by pointing out that in the presence of issuer or acquirer market power or of a subsidization of checks and cash, the socially optimal IF exceeds the level given by the tourist test. At this stage let us focus on the tourist test level as a conservative benchmark for the socially desirable IF.

B. HOW DOES THE ECONOMICS RECOMMENDATION COMPARE WITH ACTUAL POLICY?

As we noted, European regulators have endorsed the tourist test methodology.³⁹

By contrast, the methodologies proposed by American and Australian regulators are broadly similar in that they are based on the issuer’s cost. For instance, the Dodd-Frank Act prescribes an IF that is “reasonable and proportional to the cost incurred by the issuer with respect to the transaction,”⁴⁰ namely the incremental cost of authorization, clearance, and settlement. The “reasonable and proportional” phrasing allows much flexibility in the interpretation of this recommendation, making any exegesis necessarily controversial. But for the sake of the argument we can assume that it will be interpreted as an IF regulation at a level equal to the issuer’s variable cost associated with the processing the transaction.⁴¹

The issuer cost (c_i) to be used as a benchmark for the regulated IF unfortunately bears little relationship with the theoretically correct level, which focuses on the acquirer/merchant side rather than on the issuer side.

While economics tells us to take cost-based IF regulations with circumspection, it is sometimes argued that issuer cost is easier to measure than merchant benefit, which is more heterogeneous. This is probably correct and indeed, as a general point, one should be wary of policy recommendations that are based on hard-to-measure variables. This being said, merchant benefits are measurable, and there have been attempts at providing such measures.⁴² The point is that given the enormous amounts of money at stake it would be reasonable to con-

duct a couple of studies measuring benefits of specific classes of merchants and to use reasonable rules of thumb in order to extrapolate for other classes.

VI. Regulation-induced Industry-structure Distortions: Some Unintended Consequences of IF Control

Market forces are not easily suppressed and we should expect that as the dust settles on a strict IF regulation, this regulation will be evaded, possibly in inefficient ways.

A. EVADING REGULATION THROUGH MIGRATION TO THREE-PARTY SYSTEMS

A puzzle regarding the last three decades of antitrust enforcement in the payment industry is the sole focus on open systems. Such a focus tilts the industry's business model in favor of three-party systems for no clear reason. Whatever regulation (or lack thereof) one advocates, neutrality with respect to business organization should be the rule, so as to let the most efficient organizational forms emerge.

BUT THERE IS NO NEED TO QUIT
A FOUR-PARTY SYSTEM IN
ORDER TO RE-CREATE A HIGH IF
WHEN ITS LEVEL IS FORMALLY
CONSTRAINED BY REGULATION.

In reaction to downward pressure on IFs, cardholders and issuers, who benefit from higher IFs so long as merchants keep accepting the card, have an incentive to migrate toward card payment schemes that put more of the burden

on the merchant. A case in point is Australia where, in the wake of the mandated decrease in the IF, three of the top four Australian banks signed up agreements to issue American Express or Diners Club cards. IF regulation therefore induces cardholder migrations toward three-party systems that offer them a better deal in the allocation between merchants and cardholders.

Substituting merchant fee regulation for IF regulation would enable the proponents of regulation to maintain a level-playing field among competing organizational forms. But of course, this call for organizational-form neutrality does not *per se* imply that regulating three-party systems is optimal. As is often the case in second-best analysis, adding a distortion need not reduce welfare when another distortion is in place: If proposed four-party system regulation overshoots and excessively constrains the IF, the issuers' ability to evade regulation by migrating to three-party systems could be desirable.

B. PREFERRED MERCHANT PROGRAMS

In the previous regulatory evasion, issuers either joined an existing and previously closed system as licensees or started their own closed system. But there is no need to quit a four-party system in order to re-create a high IF when its level is

formally constrained by regulation. A large issuer (or a consortium of issuers) can launch a preferred merchant program. This program works as follows: The card offers low benefits and functions as an ordinary card at non-affiliated merchants; cardholders enjoy extra benefits when they shop at merchants affiliated with the program. It also signs up affiliated merchants, who then either pay a fee over and beyond the IF to the issuer for each transaction (if the issuers provide the reward), or offer direct cash rebates to cardholders, on the grounds that being part of the program brings customers to the merchant.

Preferred merchant programs are on the rise. Citi has a program called the Thank You Rewards Program that includes selected merchants and offers specific benefits tied to those participating merchants. Chase has a similar program called Ultimate Rewards. MasterCard just rolled out the MasterCard MarketPlace, which, again, offers specific rewards tied to participating merchants.

Through a preferred merchant program, an issuer, a group of issuers, or the system itself can thus increase the *effective IF*, equal to the sum of the regulated IF and the extra fee or direct cash rebate demanded from affiliated merchants.⁴³ Furthermore, the more stringent the regulation (the lower the regulated IF), the higher the resulting effective IF!⁴⁴

In a nutshell, issuers can piggyback on a regulated four-party system and, through a preferred merchant program, raise the effective IF. This probably implies some welfare losses compared to *laissez-faire* since merchants contract with (conceivably a small number of) issuers as well as the system, and consumers need to be aware of, and remember, the list of affiliated merchants (so instead of just knowing whether the merchant takes say the basic Visa card, they will need to also know whether she takes their issuer-specific premium card).

Finally, besides issuers or the system, an association of merchants may alternatively offer a card that contains reward programs. Again the multiplication of actors checking the creditworthiness of consumers may represent an inefficient bypass of IF regulation.

VII. Why the Tourist Test is Probably a Conservative Estimate for Regulatory Purposes

This section reviews two arguments suggesting that the tourist test yields a conservative estimate of the socially desirable IF.

A. ISSUER AND ACQUIRER MARKET POWER

Suppose that, as is likely, issuers make a profit at the margin on card transactions. That is, the IF that they receive from acquirers more than offsets their issuer vari-

able cost plus whatever benefit they pass through to cardholders (cash-back bonuses, frequent flyer miles, etc.). An increase in the IF, provided merchants are kept on board, boosts issuer profits.

The mandate of antitrust authorities is often interpreted as one of advocacy for consumer interests; the translation in our two-sided market context is that authorities should focus on the impact of their policies on end users (cardholders plus merchants), and thus on total user surplus.

Economists' concept of social welfare more broadly includes profits. With this broader concept of social welfare, and assuming that the acquiring sector is perfectly competitive, the internalization argument implies that the socially optimal IF is equal to its tourist test level augmented by the issuers' markup.⁴⁵

THERE IS SUBSTANTIAL DEBATE
AS TO WHETHER ANTITRUST
AUTHORITIES SHOULD FACTOR
PROFITS INTO THE COMPUTATION
OF SOCIAL WELFARE.

There is substantial debate as to whether antitrust authorities should factor profits into the computation of social welfare (they rarely do⁴⁶). Take issuer profits. If the profits associated with cardholders' installed bases are dissipated through wasteful advertising expenditures to "acquire" cardholders, profits should not enter social welfare calculations. By contrast, profits also drive technological and pricing innovations as well as new entry,⁴⁷ eventually benefitting cardholders. Indeed from a theoretical viewpoint there is a continuum of situations that vindicate various IFs at and above the narrow concept, depending on what fraction of profits are dissipated and what fraction leads to enhanced cardholder welfare. Ultimately, what fraction of profits should be factored into the computation of the IF is an empirical question, which we won't attempt to resolve here. But there is no question that not including any leads to a conservative estimate of the desirable IF.

B. SUBSIDIZED COMPETING MEANS OF PAYMENT

The analysis assumed that alternative payment methods (cash, checks) are fairly priced. This, however, need not be the case. In some countries, banks are not allowed to charge for the costs they incur on checks; in this case, checks are "subsidized" in that their cost is recovered through cross-subsidies from other banking activities. Checks and cards then wage unfair competition. Similarly, merchants may prefer cash for non-avoidable reasons (tax evasion). Cash is then unduly favored.

Basic economics teaches us that when two goods are substitutes and one of them is "subsidized," in that it does not pay some social cost it imposes on society, the other good should itself be "subsidized" so as to restore a level-playing field and prevent a wrong allocation of resources. This has long been the standard argument in favor of subsidizing public transportation to offset the unfair

advantage enjoyed by the automobile when it does not pay for its congestion or pollution cost.

The desirable policy response would be to address this regulatory failure directly by letting checks be priced and by curbing tax evasion; but if political opposition or monitoring costs make such a direct correction infeasible, then favoring card usage by raising the IF above its tourist test level is the second-best policy response.

VIII. Concluding Thoughts

Let us summarize the main insights:

A. UNDERSTANDING HOW IFS ARE SET IN THE ABSENCE OF REGULATION

- Four-party systems set their IFs, and three-party systems their merchant fees, with an eye on what merchants can bear. Because their profits grow with card volume, they have an incentive to charge high IFs/merchant fees, and to induce their cardholders to use the card.
- Merchant demand for card usage can be defined in two ways: narrow (the net benefit for the merchant: how much they directly save when a card payment substitutes for a cash or a check payment, as well as the enablement of transactions which otherwise would not occur) and broad (a concept that further includes the cardholders' perceived benefit from card usage). The narrow concept is appropriate in the case of a consumer who does not need to be attracted through card acceptance (the hypothetical "tourist"), while the broader concept applies when the merchant views her card acceptance policy as a means to attract consumers to her shop. What the merchants can bear lies between these two benchmarks.
- System competition puts downward pressure on IFs only if individual cardholders hold cards on different systems. Under full "multi-homing" merchants cannot be charged more than their net benefit.
- Three-party systems use an implicit IF, defined as the difference between the merchant discount and the acquiring cost.

B. IMPLICATIONS FOR POLICY-MAKING

- Regulated IFs should not lie below the level set by the "tourist test," which reflects the first benchmark; that is, the IF should be at least equal to the difference between the merchant's benefit from card usage and the acquiring cost.

- This level, however, probably is a conservative estimate of the socially desirable IF for two reasons:
 - ~ It does not reflect industry profit and its long-run impact on entry, innovation and end-user welfare.
 - ~ It does not reflect the negative social externalities exerted by alternative means of payment (tax evasion for cash, subsidized use for checks).
- Regulation has hitherto been misguided in that it favors closed, three-party systems over open, four-party ones. There is absolutely no economic reason for treating the two asymmetrically. Antitrust authorities should not push the industry toward a particular organizational form, but rather should let the most efficient ones emerge.

A blind application of basic economic precepts is particularly hazardous in two-sided markets. This observation however does not imply that “anything goes” in the matter of policy design. Modern economics does suggest a framework for thinking through policy-making in this area. It is my hope that this note has helped clarify the underlying principles. ▼

-
- 1 Combined, credit and debit cards accounted for \$3.5 trillion dollars worth of transactions in the United States in 2009, which is over 45% of the total purchase of goods and services. The total dollar amount of transactions has grown 29 percent since 2005. Over that same time the total purchase of goods and services increased 13 percent (Nilson Report #962). In 2009, in the United States and for debit cards alone, there were 38 billion payments, bringing in to issuers (cardholders’ banks) \$16 billion in interchange fee revenue (the interchange fee is the fee paid by the acquirer—the merchant’s bank—to the issuer—the cardholder’s bank—subsequent to a card transaction). In Europe, general purpose cards from Visa, MasterCard, American Express, and Diner Club accounted for more than \$1.8 trillion in purchase volume and another \$1.2 trillion in cash volume in 2009 (Nilson Report #950).
 - 2 The systems are not fully closed. In particular, American Express and Discover also issue their cards through some selected bank licensees (this was facilitated in the United States by a 2001 court decision saying that Visa and MasterCard could not demand exclusivity from their member banks).
 - 3 In 2008, checks still made up 26 percent of all transactions in the United States, more than in many other industrialized countries. France also had relatively high levels of check usage in 2008 at 22.1 percent (Bank of International Settlements, Country Statistics of Payment Settlements 2008). Cash was the most commonly selected payment method in 2009 accounting for one-third of all transactions in the United States (Nilson Report #962).
 - 4 Some other well-known payment systems build on top of existing ones. For instance, PayPal makes it easier for small merchants and individuals to accept cards. It charges nothing to the sender/buyer, who gives a card number, bank account number, or PayPal account number, and sets a charge for the merchant that more than covers the cost of using card systems or the cost of withdrawing from the sender’s bank account. PayPal further uses different tariff structures depending on the amount (and also on the merchant). For instance, instrumental to its acceptance for payments below \$2 (such as the purchase of a \$0.99 song) was a structure with a low-fixed (non-proportional to transaction value) amount and a high proportional levy.

- 5 We, here, do not go into the details of pricing structures (the decomposition between fixed and variable fees), which are fascinating in their own right. For example, reductions in the IF can affect cardholders' variable fees (e.g. cash back bonuses), or their fixed fee per transaction, or else (in the case of debit cards, which are linked to a bank account) the rest of the banking relationship. For an assessment of the impact of IF reduction in Australia, see DAVID S. EVANS, *ESSAYS ON THE ECONOMICS OF TWO-SIDED MARKETS*, Ch. 11 (2011), available at <https://www.competitionpolicyinternational.com/assets/Hot-Tubs/Evans-Two-Sided-Market-Essays-Final.pdf>.
- 6 For details on this point, see Robin A. Prager, Mark D. Manuszak, Elizabeth K. Kiser, & Ron Borzekowski, *Interchange Fees and Payment Card Networks: Economics, Industry Developments, and Policy Issues*, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System, (2009-03).
- 7 *Id.*
- 8 The literature has shown that, in a wide range of circumstances, surcharging deprives the IF of any role in affecting the volume of card payments. See Joshua Gans & Stephen P. King, *The Neutrality of Interchange Fees in Payment Systems*, BE PRESS, TOPICS IN ECONOMIC ANALYSIS AND POLICY, Article 1 (2003); Jean-Charles Rochet & Jean Tirole, *Cooperation among Competitors: Some Economics of Payment Card Associations*, 33(4) RAND J. ECON., 549-570 (2002); Marius Schwartz & Daniel R. Vincent, *The No Surcharge Rule and Card User Rebates*, 5(1) REV. NETWORK ECON., 72-102 (2006); and Julian Wright, *Optimal Card Payment Systems*, (47) EUR. ECON. REV., 587-12 (2003).
- 9 National Bancard Corp. v. VISA, U.S.A., 596 F. Supp. 1231 (S.D.Fla.1984).
- 10 Note that in order to regulate the IF, one must also check that the system does not undo the reduction in the IF by increasing the system fee for merchants and reducing by an equal amount the system fee paid by issuers. Also, Visa (except for Visa Europe) and MasterCard, which have moved in the last decade from a not-for-profit status to a for-profit one, can alternatively raise the acquirer system fee without lowering the issuer one.
- 11 As well as the removal of the no-surcharge rule.
- 12 Recently, competition authorities of New Zealand, Poland, and the United Kingdom have declared the multilateral setting of interchange fees illegal and to be discontinued (Prager, *supra* note 6).
- 13 Reserve Bank of Australia (2005).
- 14 See David S. Evans, Robert E. Litan, & Richard Schmalensee, *The Economic Principles for Establishing Reasonable Regulation of Debit-Card Interchange Fees that Could Improve Consumer Welfare*, Submission to the Board of Governors of the Federal Reserve System, February 22.
- 15 MasterCard has appealed to the European Court of Justice. Visa accepted to apply 0.2 percent for debit cards and 0.61 percent for credit cards for both its cross-border transactions and for some countries' domestic transactions.
- 16 See also: "...without further evidence, which MasterCard failed to submit—it cannot safely be assumed that by pursuing its member banks' aim of maximizing sales volumes MasterCard's MIF has created efficiencies that benefit all customers, including merchants" (EC MEMO/07/590, December 2007). The reader can also find useful information about the Commission's methodology in a memorandum, available at <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/09/143&format=HTML&aged=0&language=EN&guiLanguage=en>. In particular:
- As regards calculation of the (cross-border) MIF, MasterCard has engaged to apply a methodology developed in economic literature to assess efficient interchange fees which is called the 'avoided-cost test' or 'tourist test'. The fee which meets this test,

also referred to as the balancing fee, ensures that user benefits are enhanced. The balancing is such that merchants do not pay higher charges than the value of the transactional benefits that card use generates for them. Merchants derive such transactional benefits if card payments reduce their cost relative to cash payments, for instance, because transportation and security expenses for cash are saved or if check-out times at cashier desks are reduced. The implementation of the balancing fee ensures that the merchant is indifferent as to whether card or cash payments are made. To the extent that the fee is passed on to the cardholder, it will ensure that cardholders make efficient choices with respect to payment instruments, being effectively led by the MIF to internalise the cost saving that card usage entails for the merchants. Importantly, this approach prevents the MIF from being set at a level such that banks would take advantage, by collective agreement, of the fact that individual merchants feel compelled to accept a payment card even if it is more expensive than other payment instruments, fearing their customers would otherwise not make purchases at their store (e.g. because other merchants accept the card).

- 17 Neelie Kroes (European Commissioner for Competition Policy), speech/09/165, April 2009.
- 18 "The methodology underlying a MIF should be transparent to the final users of a scheme" (EC EMO/07/590, December 2007).
- 19 Chairman Bernanke famously posed this question in a Federal Reserve Open Board Meeting: "There's a presumption that prices will be set by market competition, generally, but then, of course there are counter examples such as electric utilities, for example, where the government intervention can be justified ... for various reasons. Can you ... help us thin[k] about ... what are the arguments for and against allowing interchange fees to be determined in the market versus having a regulatory intervention when we think about the economics?" See, *Federal Reserve Board of Governors Holds an Open Meeting*, CQ FINANCIAL TRANSCRIPTS, at 8 of 28 (December 16, 2010).
- 20 When issuers and/or acquirers have market power they may pass through cost increases or reductions more or less than one-for-one. Then a change in the IF may affect the price level and not only the price structure. For an analysis of pass-through in one- and two-sided markets, see the papers by Weyl & Fabinger, including E. Glen Weyl & Michal Fabinger, "Pass-Through as an Economic Tool," mimeo, Harvard University (2009).
- 21 This is more broadly a feature of "two-sided markets," see e.g. Mark Armstrong, *Competition in Two-Sided Markets*, 37(3) RAND J. ECON., 668-691 (2006); Bernard Caillaud & Bruno Jullien, *Chicken & Egg: Competition among Intermediation Service Providers*, 34(2) RAND J. ECON., 309-28, (Summer 2003); Evans, *supra* note 5, Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-sided Markets*, 1(4) J. EUR. ECON. ASS'N, 990-1029 (2003); Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37(3) RAND J. ECON., 645-667 (2006); and E. Glen Weyl, *A Price Theory of Multi-Sided Platforms*, 100(4) AMER. ECON. REV., 1642-72 (2010).
- 22 Or of surcharging for card purchases when this is allowed and does not create high transaction costs.
- 23 Even if it is priced (surcharging), imperfect information about card acceptance and about surcharging may still make the competitive outcome inefficient.
- 24 Acquiring is less competitive in countries with monopoly acquiring such as Portugal, though. And when competition in acquiring is the rule, it is less intense for small than for large merchants: smaller merchants shop around less and, furthermore, acquirers must verify their creditworthiness. See Richard Schmalensee, *Payment Systems and Interchange Fees*, 50 (2) J. INDUS. ECON., 103-122 (2002) for some implications of imperfect acquiring competition.
- 25 Jean Tirole & Jean-Charles Rochet, *Must-Take Cards: Merchant Discounts and Avoided Cost*, J. EUR. ECON. ASS'N, (forthcoming, June 2011).

- 26 If the customer instead leaves the shop, obtains some credit from his bank or a friend, and comes back to the shop to purchase the good or service, there is also a social cost, but of a different nature, since the cost is borne by the consumer and therefore fully internalized by him.
- 27 Some readers may also remember the transaction costs and delays involved when buying by phone from discount retail outlets and sending a check to pay for the goods.
- 28 A noteworthy exception is Jean-Charles Rochet & Julian Wright, *Credit Card Interchange Fees*, (34) J. BANKING & FINANCE, 1788-1797 (2010).
- 29 As shown in the literature, this assumption is vindicated under reasonable assumptions for either a non-profit association (as Visa and MasterCard were until the last decade) controlled by issuers or for a for-profit system.
- 30 More than 50 percent of American consumers in 2006 had multiple cards (that may belong to the same network, though). Most, however, made use of a single one, using the other(s) as insurance against a technical problem or non-acceptance by the merchant. That is, "multi-homing" in membership is much more prevalent than "multi-homing" in usage, see Marc Rysman, *An Empirical Analysis of Payment Card Usage*, 55(1) J. INDUS. ECON., 1-36 (2007).
- 31 See e.g. Rochet-Tirole, *supra* note 25, and especially Graeme Guthrie & Julian Wright, *Competing Payment Schemes*, 55(1) J. INDUS. ECON., 37-67 (2007).
- 32 In the case of "tourists" (as defined above), card acceptance plays no role in attracting a customer. The merchant then aims at minimizing cost and at best takes only cards that reduce her cost of transacting with the customer. When cardholders multi-home, though, the merchants turn down cards even when the latter allow them to economize on cost (the merchant fee lies below S). System competition then results in an inefficient IF strictly below the tourist test and equal to the level that maximizes expected merchant cost savings from cash substitution.
- 33 This joint surplus is called "total user surplus" in Rochet-Tirole, *supra* note 25, to which we refer for the derivations.
- 34 As explained above, American Express is a three-party system, and therefore has no formal interchange fee (the shadow IF is equal to its merchant fee minus the cost of acquiring). But the reasoning is the same as for four-party systems since the merchant is concerned about her own cost and her attractiveness, and not about the black box of the issuing and acquiring industry per se.
- 35 A well-known illustration is the Boston fee party. According to the Wikipedia American Express entry:
- However, in 1991, several restaurants in Boston started accepting and encouraging the use of Visa and MasterCard because of their far lower fees as compared to American Express' fees at the time (which were about 4% for each transaction versus around 1.2% at the time for Visa and MasterCard). A few even stopped accepting American Express credit and charge cards. The revolt, known as the "Boston Fee Party" in reference to the Boston Tea Party, quickly spread nationwide to over 250 restaurants across the United States, including restaurants in other cities such as New York City, Chicago, and Los Angeles. In response, American Express decided to reduce its discount rate gradually to compete more effectively and add new merchants to its network such as supermarkets and drugstores. Many elements of the exclusive acceptance program were also phased out so American Express could effectively encourage businesses to add American Express cards to their existing list of payment options.
- 36 See Rochet & Tirole, *supra* note 25. An early analysis of unobserved merchant heterogeneity is due to Schmalensee, *supra* note 24.

37 Rochet & Tirole, *supra* note 8.

38 John Vickers, Public Policy and the Invisible Price: Competition Law, Regulation and the Interchange Fee, 231-247, Proceedings of a conference on "Interchange Fees in Credit and Debit Card Industries," Federal Reserve Bank of Kansas-City, (May 4-6, 2005).

39 In a European Commission memorandum, it adds some caveats in answering the question, "will any MIF that satisfies the 'tourist test' be automatically compliant with Article 81 (3) EC Treaty?":

The 'tourist test' provides a reasonable benchmark for assessing a MIF level that generates benefits to merchants and final consumers. It determines a MIF that allows the promotion of efficient payment instruments, while at the same time preventing that the MIF exploits business-stealing effects to the detriment of the scheme's users, which would lead to an inefficient promotion of payment instruments that impose invisible costs on consumers. However, the general applicability of the 'tourist test' for the purposes of Article 81 (3) depends on the specifics of the markets at hand. Some (non-exhaustive) cautionary examples are listed below:

1. While a MIF at appropriate levels makes the use of efficient payment instruments more attractive to consumers, other (less-restrictive) mechanisms may do so as well in some markets. For instance, this is the case if merchants themselves can be expected to efficiently incentivize the use of less costly payment instruments by applying rebates to those means of payment. In this case a MIF may not be indispensable, as direct incentives given by merchants may internalize network externalities between merchants and users of payment instruments more directly.

2. When a payment card would reach universal usage in a market even without MIF, the need to promote the issuing of such a card in terms of network effects would vanish.

3. More generally, there must be a reasonable channel through which interchange fees can promote the use of cards. With respect to debit cards, the reward programs for such cards (which directly incentivise usage) typically do not exist and that card-holding across Member States is already widespread (but not complete). Therefore, the DG Competition does not consider that possible future increases of the 'tourist test' estimation for debit cards would necessarily justify an increase in the debit card MIF, unless payment card associations can ensure that the banks receiving such a higher MIF have installed appropriate cash-back programs for debit cards that could directly incentivise a wider use of debit cards on a per-transaction basis.

4. Conversely, circumstances may in principle arise under which justifications for higher MIFs could be demonstrated by payment card associations. However, significant objective evidence would be needed to establish that this is the case.

Antitrust: Commissioner Kroes notes MasterCard's decision to cut cross-border Multilateral Interchange Fees (MIFs) and to repeal recent scheme fee increases—frequently asked questions, EC Memo/09/143 (2009), available at, <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/09/143&format=HTML&aged=0&language=EN&guiLanguage=en>

40 See new section 920 of the Electronic Fund Transfer Act.

41 In its December 13, 2010 recommendation, the Board of Governors of the Federal Reserve System's staff recommends using an average cost measure as a means to calculate incremental cost.

42 See Daniel Garcia-Swartz, Robert Hahn, & Anne Layne-Farrar, *The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics* 5(2) REV. NETWORK ECON., 175-198 (2006); Daniel Garcia-Swartz, Robert Hahn, & Anne Layne-Farrar, *The Move Toward a Cashless Society: Calculating the Costs and Benefits* 5(2) REV. NETWORK ECON., 199-228 (2006); and Anne Layne-Farrar, *Assessing Retailers' Costs and Benefits from Accepting Debit Cards*, (2011). Also, "The Commission's competi-

tion department has commissioned a study with a view to collect data in order to improve the factual basis for the assessment of what level of MIF would be in accordance with the tourist test," European Memorandum, *supra* note 39.

- 43 I here assume that cardholders are aware of merchant card acceptance policies. If regulation is more effective if cardholders are "tourists," but we know then that regulation is always dominated by *laissez-faire* in that case.
- 44 A more stringent regulation lowers total user surplus and makes the card less appealing to the merchant when considering attracting consumers. This makes the merchant more likely to accept high effective IFs. Because it maximizes joint user surplus, the tourist test level is not subject to such arbitrage by preferred merchant programs.
- 45 As usual, the existence of an issuer markup above issuing cost does not imply that the issuing industry is not competitive from a long-run perspective. It may be that the markups cover the fixed costs associated with the issuing activity.
- 46 In this respect, competition policy takes a somewhat different approach from intellectual property law, which views patents and other profit-generating IP protection institutions as an inefficient, but key instrument for providing incentives for innovation.
- 47 Note that we here take a long-term perspective. In the short-term, profits are just rents that go to investors. Accounting for profits raises a different interrogation when one takes the short-term perspective; if investors are average citizens they should be fully accounted for. With well-to-do investors and redistributive concerns, only a share of profits should be included into social welfare.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Innovation Market Theory and Practice: An Analysis and Proposal for Reform

Kent Bernard

Fordham Univ. School of
Law

Innovation Market Theory and Practice: An Analysis and Proposal for Reform

*Kent Bernard**

Encouraging and/or preserving innovation in mergers and acquisitions have been critical factors in modern antitrust analysis. These aims have been justification for the breakup of proposed research programs targeting diseases as serious as HIV/AIDS and cancer. The rationale given is always to protect competition and enhance the benefits to consumers.

Lawyers and economists justify intervention in mergers based on predictions of what will or might happen many years down the road in scientific research programs. They base those predictions on various theories and assumptions of how companies behave. But an examination of the actual drivers in the research-based pharmaceutical industry, such as the time factor of revenue destruction and the resulting continuing need for new products, along with a review of what happened in key cases after the agencies acted, reveals that those underlying assumptions may well have been unfounded.

This factual consideration of how business actually behaves has been missing from the analysis. This article looks at the leading approaches to “innovation markets.” It then reviews the key cases in which the theory has been applied, and looks to see what actually happened after the case files were closed. In other words, did the intervention do any good, and/or did the lack of intervention do any harm?

*Adjunct Professor, Fordham University School of Law, JD 1975 University of Pennsylvania; BA 1972 Colgate University. I would like to thank Commissioner Thomas Rosch (U.S. Federal Trade Commission); Professor D. Daniel Sokol (University of Florida); Paul Lugard (Royal Philips Electronics and Fellow at Tilburg University Law and Economics Center, Netherlands); Luc Gyselen (Arnold & Porter, Brussels), Lindsay McSweeney, and Marcus Glader (Vinge Law Firm, Brussels) for their comments on earlier drafts of this paper. Any errors or oversights are entirely my responsibility.

The results of that inquiry strongly suggest that not only was the intervention not beneficial, it may have dampened innovation by reducing the potential reward while ignoring the risks that any innovator is being asked to run.

Innovation market theory arose out of a concern that mergers were reducing innovative capacity. The regular tools of analysis failed to provide a remedy for this sort of highly speculative harm, so the agencies stretched the concept of innovation markets to allow them to act under it. However, the analysis here shows that the perceived risk was based on a misapprehension about how companies actually behave and the nature of innovation itself. Once that is understood, the need to stretch the concept of innovation markets goes away.

This article also proposes an alternative approach, grounded in traditional antitrust but based on market reality rather than theory. When this approach is applied to the facts of the cases, it allows intervention when needed while avoiding speculative interference with scientific and business pursuits.

I. Setting the Stage

The concept of preserving competition in what have been classified as “innovation markets” has been remarkably resilient. It has been the justification for the breakup of proposed research programs targeting diseases as serious as HIV/AIDS and cancer. The rationale given is always to protect competition and enhance the benefits to consumers. But as U.S. Federal Trade Commissioner (“FTC”) Tom Rosch noted, arguing over whether the parties to a merger have market power in an innovation market is a bit like trying to fit a square peg into a round hole. Those markets just can’t be pinned down under traditional antitrust concepts.¹

When one digs a bit below the surface of the innovation market concept, it becomes more and more difficult to figure out whether the application of that concept in an antitrust case has led to good results or bad ones. The actual basis for defining an innovation market in a given case is almost impossible to pin

ARGUING OVER WHETHER
THE PARTIES TO A MERGER HAVE
MARKET POWER IN AN INNOVATION
MARKET IS A BIT LIKE
TRYING TO FIT A SQUARE PEG
INTO A ROUND HOLE.

down. As will be discussed further below, what we are seeing is often a future goods analysis, divorced from its normal limits in terms of timing and likelihood of market entrance and being extended beyond its limits by cloaking it in innovation market language.

What has been missing from the analysis is a consideration of how business actually behaves, which I believe should be the starting point in any decision whether to intervene in a transaction. This article will look at the leading approaches to innovation markets, dissect what they mean, and look at what they intend to accomplish. It will then revisit the key cases in which the theory has been applied (nearly all of which involve pharmaceutical research and development) to see what actually happened after the case files were closed. In other words, did the intervention do any good, and/or did the lack of intervention do any harm?²

Lawyers and economists now second guess scientists and business people in terms of predicting what will or might happen many years down the road in scientific research programs. But while various theories and assumptions tell how companies should behave and are used to construct rationales for intervention, an examination of the actual drivers in the research-based pharmaceutical industry demonstrates that many of those assumptions are not correct. And to the extent that those assumptions are what underlie the intervention, then the intervention is unsupported.

In these unsupported cases, intervention can seldom be shown to have increased or preserved innovation in the sense of leading to more or quicker products to market. Indeed, results after the cases have been resolved raise serious questions whether such intervention dampened innovation by reducing

potential rewards while ignoring the risks that any innovator must run to be a successful market participant.

This article proposes an alternative approach, grounded in traditional antitrust but based on market reality rather than theory. Just as the end point of innovation is a tangible outcome, the definition of innovation markets needs to be tied to something tangible as well. The 1995 Intellectual Property Guidelines (the “IP Guidelines”) made a strong connection between innovation and something that can be grasped, owned, or measured. It limited innovation market inquiries to cases where the parties have unique access to necessary **tangible** assets and where the capability to engage in the relevant research and development (“R&D”) can be associated with specialized assets or characteristics of specific firms.³ Reviewing the approaches that came after those Guidelines, and measuring them against what actually has taken place and how business actually behaves, leads to the conclusion that this modest definition from 1995 provides the best real world anchor for the theory, allowing intervention when needed while minimizing purely speculative interference with scientific and business pursuits.

JUST AS THE END POINT OF INNOVATION IS A TANGIBLE OUTCOME, THE DEFINITION OF INNOVATION MARKETS NEEDS TO BE TIED TO SOMETHING TANGIBLE AS WELL.

II. The Prehistory of Innovation Market Theory

In its most obvious meaning, an innovation market would mean a market for innovation itself, suggesting the auctioning off of a team of expert scientists who are the only ones in their field producing the result that if one bidder wins, everyone else loses. That clearly is not factually accurate. Any workable theory needs to come up with a more useful and practical approach.⁴ And no matter how creative agencies may want to get, at the end of the day any analysis of markets is tethered to statutes and regulations, Section 7 of the Clayton Act in the United States, and Article 101 of the TEUF (and the Merger Control Regulation in the European Union). Unless the defined innovation market is at least consistent with the statutes and precedents, it is not much more than an interesting academic exercise.

Perhaps the most striking characteristic of the early cases cited for the development of innovation markets is that those cases, on their facts, did not need to speak of “innovation markets” at all. The concept was thrown in, but neither the facts nor the holdings required it.

For example, *Smog Control Devices (1969)*⁵ was an alleged agreement among car manufacturers to slow down development of pollution control devices and make sure that no one car maker got ahead of another. As a horizontal agreement not to compete in a field, no new kind of analysis was required to condemn it.

U.S v. GM (1993)⁶ involved certain truck transmission production facilities that were characterized as a specialized asset. The innovation market consisted of the two companies with distinctive assets in place to do R&D, manufacturing, and sales in a limited and defined product market with high entry barriers. While the Department of Justice (“DOJ”) tried to claim that the case was about a broader innovation market,⁷ on the facts of the case there was no need for any kind of new theory.

Rereading some of the material from 1990-2000, one comes away with the strong sense of *déjà vu*; that the enforcement agencies were trying to create a broader rule by adding language to cases where no broader reach was required by the facts, and then talking about the broader rule as if it was established law.⁸

PERHAPS THE MOST STRIKING
CHARACTERISTIC OF
THE EARLY CASES CITED
FOR THE DEVELOPMENT OF
INNOVATION MARKETS IS THAT
THOSE CASES, ON THEIR FACTS,
DID NOT NEED TO SPEAK OF
“INNOVATION MARKETS” AT ALL.

The next major development in innovation market analysis came with the publication by the DOJ and the FTC of the IP Guidelines in 1995. In discussing the markets that could be affected by licensing arrangements, the Guidelines broke down the universe into three types of markets: (1) Goods; (2) Technology (licensing); and (3) Innovation/R&D.⁹

The IP Guidelines recognized that Innovation or R&D presented different issues than markets made up of goods or technology, and that an unchecked definition of innovation markets could lead to unguided intervention. Indeed, this is what seems to have taken place in some cases. Why this is so, and what it has meant for innovation in the real world, will be discussed below.

III. Why Is Innovation Important?

Before analyzing how to best define innovation, it would be good to explore why that question is important. Start with classic paradigm of the white-coated person in a lab. Why are his actions of importance to anyone else?

First, of course, it can be good to extend the thresholds of knowledge for its own sake. Also, smart people like/need to have time to just explore areas in order to keep their minds sharp for more commercially dedicated disputes.

But the main reason that people care about research or innovation is because it can lead to new or improved products (and processes) in the future.¹⁰ This may result in making existing products better and/or less expensive for consumers, or the development of new products, such as more efficient power sources, cleaner air, or new medicines to treat diseases. And this leads us to a point that tends to get overlooked in the debate. R&D has value, in large part, because the end point has value. And that end point can almost always be measured in a product market.

FTC Commissioner Rosch focuses on exactly this point, when he defines the key question when analyzing innovative market actions as “[W]hether from a policy standpoint, the application of antitrust laws to innovation markets provides consumers with better products or products that are developed more quickly.”¹¹

So the question becomes how the DOJ or the FTC can predict today, when the decision whether to intervene in a transaction has to be made, what will be the results of given R&D—if and when it leads to any results at all. This often is a very fact dependent analysis. Society may well be better off in some cases having two or three projects in the hands of one company rather than in three separate companies (where that one company has the scientists, the money, and the infrastructure to bring the research to fruition as one or more products, whereas other companies are too small/thinly funded/scientifically light to advance the projects).¹² This is not to say that this is always the case. It certainly does appear that the question is fact dependent.¹³

But before an attempt can be made to analyze any particular real world fact situation, there are a couple more awkward questions for any innovation markets theory or theorist:

1. How can one determine how much R&D is good, or better?
2. Can someone monopolize the R&D that has been so identified and, if so, how?

IV. How Does One Measure, Acquire, or Monopolize Research and Development?

How does one measure innovation? Make it more concrete: how does one determine how much R&D is “enough” or “right” or “too little”? These terms only make sense within a system that allows measurement.¹⁴ So here are some possible measures of R&D:

1. Amount of money spent;¹⁵
2. Number of patents;¹⁶
3. Number of products in development, or launched.¹⁷

None of these seems really satisfactory. What is missing is a measuring rod, and then some kind of boundary condition (to determine what is being measured). If the standard is the number of patents, for example, one needs to ask, “patents for what?” This is really simply another way to revisit the matter of defining an innovation market—how can one define it when there is no product yet and perhaps never will be?

So should R&D be measured by number of compounds or products in development? At what stage? To try to measure R&D by spending costs highlights the fact that not all spending is effective. To try to measure R&D by the number of compounds or projects simply encourages odd counting and measuring for the sake of measuring. If a company is studying one compound for three uses, is that one or three in the measuring system?¹⁸

Such a simplistic counting cannot be enough. If one company has five compounds in research for treating bacterial infections, and another company has five compounds in research for treating high blood pressure, this says very little about what a merger would do. Putting the projects together would not seem to lessen any work in either field. And even if an enforcer could do something with the numbers internal to the merging parties, it would still need to know who else is capable of and/or is doing work in either field before that enforcer could figure out what the numbers meant.

The theoretical analysis keeps crashing on one basic rock—to monopolize or to reduce competition, there has to be a defined market.¹⁹ Effects do not take place in the ether.²⁰ So let's take a different tack for a moment.

Is the concern about a reduction in the number of projects in a field, or really about a reduction in the independent innovative capacity in that field? If the answer is “the number of projects,” then you need to explain how you determine an optimal number of such projects. That is heavily dependent on the facts of each case. More projects may be better than fewer, but more projects also may be worse (three weak candidates may not be better than one strong one).

THIS VIEW MAKES IT INHERENTLY
DIFFICULT TO IMAGINE ANYONE
MONOPOLIZING R&D
IN ANY SENSE OR IN ANY FIELD.

So what about independent innovative capacity—could someone monopolize it, and what would that mean?²¹ While patented technology can be monopolized, the components of modern R&D (scientists, laboratories, computer access) are available worldwide.

This view makes it inherently difficult to imagine anyone monopolizing R&D in any sense or in any field. Perhaps for this reason the 1995 IP Guidelines came at the issue from the flank. They limited innovation market inquiries to cases where the parties have unique access to necessary **tangible** assets; where the capability to engage in the relevant R&D can be associated with specialized assets or characteristics of specific firms.²²

This approach clearly would work in the Smog and the truck transmission cases. But what constitutes a “specific asset” isn't always obvious. Back in 1995 Richard Rapp raised the concern that the agencies would simply ignore the

“specialized assets” requirement,²³ and the cases that have followed suggest that may have been exactly what has happened.

Most innovation market cases are in the pharmaceutical field, and almost all are settled by consent order. A company may agree to a divestiture because the alternative is significant delay in getting the deal done. And given the odds against success for any given project, a fight to death to save one R&D project may well not be worth having.²⁴ But the fact that the merging parties may have given up on an issue does not mean that intervention was justified, correct, or helpful.

V. Why Should Society Worry about Research and Development Projects, and What Should the Goal Be?

Society cares about research and development, in fact in innovation in general, because it can lead to new or improved products (and processes) in the future. These improvements may result in making existing products less expensive for consumers or the development of new products—whether that means more efficient power sources, cleaner air, or new medicines to treat diseases. R&D has value, in large part, because the end point has value. And that end point can almost always be measured in a product market. In other words, I am looking to regulate the inputs based on a hypothetical impact on the outputs.

Acquiring research and development or innovative capacity is clearly different from acquiring something such as a raw material source. The kind of innovation being discussed here requires access to scientists and other people, so surely whatever it is being spoken of as being “monopolized” cannot be controlled in the same sense that one can monopolize a market for garbage collection by purchasing all of the outstanding permits in a town or city.²⁵ Much of the discussion about innovation and research speaks in terms of what might be under various scenarios. But while these theoretical constructs are often ingenious and sometimes elegant, they often fail when one looks at actual cases and analyze what has happened after either the intervention or non-intervention of the authorities. What needs to be done is to take the argument from “what might be” down to “what is.”

Perhaps the most ambitious recent attempt to grapple with this area is Michael Carrier’s, who deals with potential relationships between market structure and innovation, and constructs an ingenious test based on various theories of innovation suppression and competitive activity.²⁶ There is much

BUT WHEN THE ARGUMENT
MOVES TO DISCUSSING THEORIES
OF WHETHER OR WHY A MERGING
FIRM MIGHT SUPPRESS
INNOVATION, THE ANALYSIS
UNFORTUNATELY DOES NOT
REFLECT THE REALITY OF THE
CURRENT RESEARCH-BASED
PHARMACEUTICAL INDUSTRY.

valuable material in his discussion of pharmaceutical R&D cases, and his frank approach at looking at compounds reasonably likely to make it to market. After analyzing the data, Carrier defines “reasonably likely” for pharmaceutical R&D as Phase III (where the chance of success is over 50 percent and the timing is 2-4 years).²⁷ This is the same standard routinely used by the European Commission in such cases.²⁸ Phase III compounds are real future goods.

But when the argument moves to discussing theories of whether or why a merging firm might suppress innovation, the analysis unfortunately does not reflect the reality of the current research-based pharmaceutical industry.

VI. The Reality of the Research-Based Pharmaceutical Industry

The unceasing need to generate new products and new revenues, the uncertainties of R&D, and the FDA’s approval process and timing all strongly argue against any assumption that a company would try to retard innovation by acquiring a company and then suppress its R&D. One counter hypothetical is often given, but it actually supports the point. In the situation where one company has a dominant product on the market and the other company has the late stage compound most likely to disrupt the market during the patent life of the existing product, a classical “actual goods”/“future goods” analysis counsels one to look closely at the transaction. But this is not an innovation market scenario and does not impact innovation *per se*.

The critical point for the research-based pharmaceutical industry that often is overlooked is patent life. Any monopoly that may result from patent protection has a defined life and a defined end point. This life span needs to be a key part of any analysis of what parties are likely to do.

This industry depends on patents to an extraordinary extent.²⁹ And in the drug field, patents provide a shorter effective life than in almost any other field as a result of the long testing process that has to take place before a patented compound can become a marketed drug product.³⁰ When that realization is combined with the fact that the vast bulk of the expenditures in drug R&D are loaded into Phase III (the large scale clinical tests³¹), and that even there over 40 percent of the compounds fail, you have a context where finding the next successful compound is a never ending hunt.

However, this context had not led to the extinction of “small science” (companies of less than enormous size or what used to be called small- or mid-sized companies). In fact, it has led to an interesting multi-tier structure, with large companies that can and do oversee broad scale clinical testing (the “Development” in R&D), and a large number of smaller companies (some much smaller) that do basic research. Many of these smaller companies are funded by

venture capital firms, which provide money up front hoping to cash out if the science is successful and the company can be sold to a large pharmaceutical company or the product licensed out on good terms.³² In addition, scientific research is done in countless universities, many of which have made substantial sums licensing their results out to pharmaceutical companies. Perhaps the most famous example is the Cohen/Boyer patent on cloning at the University of California at Los Angeles, which earned the university over \$300 million in license fees and royalties.³³

So, to say as Carrier does, that the pharmaceutical industry meets the test for applying innovation market analysis because “the capabilities to engage in the relevant [R&D] can be associated with specialized assets or characteristics of specific firms,”³⁴ is an understandable attempt to create an analytical framework, but ultimately is either tautological (these are the only firms that can do the work because these are the only firms doing the work) or not in accord with reality. Indeed, over time, companies that have worked in one disease area often shift to another. It does not mean that they were incapable of doing work in the second disease area before, only that they chose not to do so.

Assets are always limited, and the allocation of assets (including research spending and direction) is a key function of management. Even a company investing billions of dollars cannot be invested in every potential disease area and scientific approach. However, assets can and have been reallocated. To look at current activity and conclude that everyone not in a certain field must be incapable of working there is to jump to an unsupported conclusion.

TO LOOK AT CURRENT ACTIVITY
AND CONCLUDE THAT EVERYONE
NOT IN A CERTAIN FIELD
MUST BE INCAPABLE OF WORKING
THERE IS TO JUMP TO AN
UNSUPPORTED CONCLUSION.

The hunger of big pharmaceutical companies for new drugs is insatiable.³⁵ There are three reasons for this. First, finding, developing, testing, and selling drugs are what drug companies must do to continue to exist. Second, once a company has reached a level of sales, it needs to stay there (or increase it, along with profits) to satisfy its shareholders. Third, products are not static. In the prescription drug universe there is no such thing as having a “natural monopoly” that can continue indefinitely. Once a major drug loses patent protection, generic versions quickly come on the market and drive the price down dramatically.³⁶ Indeed, sometimes a company’s sales can be hurt when someone else’s drug goes generic (and therefore becomes cheaper and the preferred choice of payers such as governments and insurance companies).³⁷

So while economic theory might counsel that a company “should” sit back and milk the “monopoly” cash cow, the realities of the pharmaceutical market place impose a different paradigm. It is the *time factor of revenue destruction* that is often omitted from the analysis, but which, in fact, drives the business decisions.

Consider the following hypothetical case. A company has a major prescription drug product on the market, with seven years left on its key patent. This company sees a compound that is just entering Phase III that shows great promise in that same field and is available for acquisition. Should the operating assumption for the DOJ or FTC be that the company making the acquisition would develop the compound or suppress it?

Based on the market realities discussed above, in almost every case the company will want to develop that new product, for some fairly evident reasons. The company's existing product has a limited financial life, and that time is running. And there may be other products that compete with it that are going off patent sooner, which will add even more pressure on the company's product. Even a Phase III pharmaceutical compound has a 40 plus percent chance (on average)

IT IS THE TIME FACTOR OF
REVENUE DESTRUCTION.

of failing. The company needs one or more new products to pick up the slack when the revenue stream from the old one dies.

Not only can a company have more than one drug in a field (i.e. potentially competing products, differentiated in marketing), it likely wants to have another product on the market before the patent expires on its existing one, so that it can move prescribers to its new (patent protected) product. And, of course, based on the failure rates of compounds, a smart company will want to have multiple candidates in the pipe line, in case one crashes late in the game. It is depressingly easy to find examples of such late stage failures.³⁸

If the company buys and suppresses the new compound, when the patent on its existing product expires, the company has nothing. The thought of losing a major revenue stream and having nothing to replace it can, and should, give management nightmares. The idea that a company would buy up potential next generation products in order to kill them off simply does not accord with reality in the research drug industry. In fact, one could reasonably argue that the company already in a market has at least as great an incentive to develop the next generation product (or develop a compound acquired from outside) than does any other company.³⁹

In much of the analysis there seems to be an underlying unexpressed bias that society would be better off if each compound was owned by a separate company. This atomistic model is not supported by any research of which the writer is aware. And it is contradicted by the fact that people working in a field often become better in that field over time. A company working on AIDS drugs is more likely to develop the next drug than a company that has never worked in the area.⁴⁰

VII. Clearing a Path to a New Theory of Innovation Markets

Before one can make a sensible proposal for how to handle innovation markets, it is necessary to set out just what would qualify as such a market under the definition.

A. INNOVATION MARKETS SHOULD BE A LAST RESORT ANALYSIS

Given the problems in defining innovation and determining which conditions help or hinder it, innovation markets should not be the first choice to use a context for analytical approach. If something fits under a more solid and established category, that category should be used.

1. Any subject area in which there is a product already launched should not be treated as an innovation market. It can be treated as an actual goods/future goods market, with which the enforcement agencies have a lot of experience. Recall that the whole point of innovation is to create and produce new products. In the prescription pharmaceutical area, Phase III compounds and/or anything likely to be approved within about 2-3 years should qualify as an initial screen.⁴¹ This future goods/products idea is the general approach taken by the European Commission in the proposed reform of the guidelines for cooperation among rivals.⁴²
2. Where the market consists of IP, this should be analyzed as a property market and not an innovation one. If one company owns a portfolio of patents in a field and the merging partner owns a complementary portfolio, combining them may preclude others from doing research, or at least make it more expensive to do so. But this has nothing to do with the idea of innovation itself. Patents are assets.⁴³ If one company has such assets in a field, and it attempts to acquire more of those assets, the competitive effects of this acquisition can be analyzed using traditional antitrust theory.

These alternative approaches should be applied to many cases formerly classified as innovation market cases. Of those cases that remain (i.e. outside of the actual goods/future goods or IP markets), I will try to see how they can be analyzed in terms of what potential harm would be allowed by the merger, whether that potential harm is likely or plausible under real world conditions (based on what actually happened), and whether such potential harm is likely enough to occur to justify intervening in the transaction.⁴⁴

The analysis starts from the premise that the parties should be allowed to make their bets (after all, a merger is actually a bet that the two companies can operate more efficiently as one than they did as two) without interference from antitrust agencies, unless the agencies can show a real potentially adverse impact on competition.⁴⁵

Also, while it is seductive to think that the peculiarities of pharmaceutical regulation and R&D can mean that the capabilities to engage in the relevant [R&D] can be associated with specialized assets or characteristics of a small number of specific firms,⁴⁶ a quick look at investments by venture capital firms will reveal that there are dozens, if not hundreds, of small inventors in the drug

THE ONLY SPECIALIZED
CHARACTERISTIC THAT COMPANIES
NEED TO HAVE IN ORDER TO DO
PHARMACEUTICAL RESEARCH AND
DEVELOPMENT IS WEALTH AND THE
WILLINGNESS TO PLACE LARGE
BETS ON SCIENTIFIC CANDIDATES
THAT MIGHT NEVER BECOME
SUCCESSFUL PRODUCTS.

industry. There also are countless universities, all eager to partner with companies.⁴⁷ The only specialized characteristic that companies need to have in order to do pharmaceutical research and development is wealth and the willingness to place large bets on scientific candidates that might never become successful products. But by that standard, surely Goldman Sachs qualifies as one of the potential participants, as does Exxon-Mobil.

So with this as prelude, it is time to look at some key cases, and see what remedies were ordered (or why they were not), and what actually happened after Dorothy went back to Kansas and the case files were closed.⁴⁸

B. WHAT THE AGENCIES DID, AND WHAT HAPPENED NEXT

1. 1990—Roche/Genentech⁴⁹

The FTC alleged a market to be “CD4-based therapeutics for the treatment of AIDS and HIV infections.” The allegation was that a limited number of companies were developing CD4 based therapy, and that Roche had patent applications pending on its compound (but not on the field as a whole, so as to preclude anyone from doing work).

Even assuming that isolating a type of attack on a disease is a legitimate way to define a market (the analysis does not pivot on this point), Genentech was in Phase I studies of its compound, and Roche had not even entered the clinic with its compound. A third company, Biogen, was in Phase I/II studies with its compound.⁵⁰ If this is a product market, and the FTC is looking at future goods, these companies are too far away from market production and the odds against success are too great to warrant intervention. Recall that a Phase I compound has only a 10-15 percent chance of reaching the market, and likely will take 8-10 years to do so.⁵¹ A pre-clinical compound is even farther back than that, with an even lower rate of success.

Roche was required to grant non-exclusive patent licenses to its technology. All of the projects later failed.⁵²

Whatever the merits of a product market approach here, an innovation market attack fails at the start. While only a limited number of companies were

developing CD4 based therapies, there was no allegation that there was any practical limit on the number of companies that could undertake such a project. In the terms of the IP Guidelines at 3.2.3 there is no reason to think that “the capabilities to engage in the relevant [R&D] can be associated with specialized assets or characteristics of specific firms.” In fact, this approach to treating AIDS simply was a high-risk proposition approach that most firms chose not to take. And based on the results, those other firms were right. What the FTC did was to take what should have been an analysis based on future goods rules, apply it to compounds that were very far removed from reaching the market, and wrap the analysis up in innovation market language.

By intervening, the FTC, in effect, told the parties that it was reducing the potential rewards from pursuing a risky and expensive research venture (and one targeting a serious health issue—AIDS), in order to make sure that in case the parties did succeed, another party might be able to copy the same approach. That intervention was potentially harmful and, at best, not helpful.

BY INTERVENING, THE FTC, IN EFFECT, TOLD THE PARTIES THAT IT WAS REDUCING THE POTENTIAL REWARDS FROM PURSUING A RISKY AND EXPENSIVE RESEARCH VENTURE (AND ONE TARGETING A SERIOUS HEALTH ISSUE—AIDS), IN ORDER TO MAKE SURE THAT IN CASE THE PARTIES DID SUCCEED, ANOTHER PARTY MIGHT BE ABLE TO COPY THE SAME APPROACH.

2. 1995—*American Home Products/American Cyanamid*⁵³

The alleged market was a vaccine to treat rotavirus. No such product existed. The allegation was that the merging companies were two of the three producers with projects either at or near the clinical trial stage of FDA review. In fact, American Home Products (“AHP”) was in Phase II/III studies, and American Cyanamid appeared to be still preclinical.⁵⁴ The FTC required that the American Cyanamid project be licensed out.

In terms of future goods, the American Cyanamid project clearly was too far out to be any sort of a factor. It was at the preclinical phase, which gives it less than a 10 percent chance of success and a time line of likely at least 8-10 years.

One can make a powerful argument that no intervention would have been the best course. If the AHP compound failed, which it eventually did in 1999 based on a side effect (after FDA approval and launch), then AHP could have applied that knowledge to the other project. And if the AHP product succeeded, it was so far ahead of the time line for the Cyanamid compound that there well might have been no market overlap at all. In fact, the Cyanamid compound never reached the U.S. market.⁵⁵ Another company entirely, Merck, launched its own rotavirus vaccine in 2006.

3. 1995—*Glaxo/Wellcome*⁵⁶

The alleged market was a specific chemical approach to treating migraine headaches (using 5HT-1D agonists). Each party was developing an oral form of such a drug. Glaxo had a product on the market in injectable form. Glaxo also had an oral Phase II/III compound. Wellcome had a Phase III compound, likely to be the first to market for the oral form. The FTC required the divestiture of the Wellcome Phase III compound.

If the market is defined as treatment for migraine, or even 5HT-1D agonists for treatment of migraine, then there exists an actual product market with one party as the dominant seller and the other party with a Phase III compound that is the most likely next entrant (future good). There is no need to talk about innovation markets at all on these facts.

But what if one looks at injection as a disfavored method of administration, so that there are no existing products but just two research programs? In fact, although this was not included in the case data, it appears that Glaxo was working on spray and tablet versions of its injectable product, and these were approved by the FDA in 1997.⁵⁷ Glaxo continued to lead the market through at least 2006.⁵⁸

As an actual goods/future goods case, this is straightforward. The remedy would be justified, and there is no reason to get to innovation markets at all.

4. 1997—*Ciba/Sandoz*⁵⁹

The subject was gene therapy products and research. No one had a product on the market, but the merging parties were alleged to control the IP necessary to commercialize products in the field. They were also identified as two out of only a few entities capable of commercially developing such products, but it is unclear if this was a separate allegation towards an innovation market theory or simply a restatement of the IP position. If the companies controlled the key IP, then they could exclude others from the field and therefore, by default, they were among the few (if not the only) ones legally capable of doing work in the field. Various non-exclusive licenses were required to allow the merger to go forward.

As of 2008, there was no gene therapy product on the market. In contrast to the forced divestitures of compounds or R&D projects, the issue here is more clearly viewed as one of an IP market—the market was the IP allowing/preventing others from doing work in the field. Allowing the companies to merge potentially created a patent bar that would not have existed but for the merger. So even if many companies had the scientific ability to do R&D in the field, they would not have been able to because of the IP block. One would have to know what the patents covered, and what was covered by the required licenses, to make a full evaluation.

5. 2000—Pfizer/Warner Lambert⁶⁰

While a number of issues could be raised about this case, the innovation market issue is framed by the FTC's definition of an innovation market consisting of research and development of epidermal growth factor tyrosine kinase inhibitors (referred to as EGFR-TK inhibitors) for the treatment of solid cancerous tumors. There are various ways to treat such tumors. As for the one at issue,

“While the complete mechanism of action is not entirely understood, the drug appears to impede cell-cell signaling pathways which have been implicated in rapid cell division and survival. Over activation of these pathways are thought to be central to tumor growth and metastasis.”⁶¹

This quote is significant because it makes clear that however this compound may work, it is only one of a number of approaches to blocking tumor growth. At the time of the transaction, there was no EGFR-TKi product on the market. AstraZeneca had a Phase III compound, Imclone had a Phase III compound, Pfizer had a Phase II compound (in a partnership with OSI, a small biotech company), and Warner Lambert had a Phase I compound that arguably used a different mechanism of action. So even on this market definition, there were four companies in the market, the merging parties were the farthest behind, and no one suggested any limit on the number of companies that could do work in the field (and might well do so if the concept proved to be effective and safe).

On its face, given that two other companies were more advanced even in the limited field being considered, and that the merging parties were in relatively early stage development, it is hard to see how intervention was justified. The FTC required the divestiture of the Pfizer/OSI compound (the more advanced one), likely because the partner OSI could be relied upon to continue the work with less potential uncertainty as would have existed with an unrelated purchaser of the Phase I compound. Indeed, OSI did more than that. It partnered first with Roche and Genentech for \$187 million⁶² and in 2010 the entire company was sold to Astellis for \$4 billion, in large part on the performance of the compound.⁶³ The compound that Pfizer was allowed to retain never got out of the testing phase.

The end result was three products on the market using the designated pathway: Imclone, OSI, and Amgen (not even on the charts in 2000). AstraZeneca's product was put on the market, but pulled in 2004 for lack of efficacy.⁶⁴

Even on the very narrow market definition, out of two Phase III compounds, one made it to market and stayed; out of two earlier stage compounds, one made

it to market. The facts following the merger are solid: there would have been no harm to competition if no divestiture had been required.

It is worth exploring the language of the Order for what it tells us about the FTC's express analytical process. First, the FTC claimed that Pfizer could delay one compound or drop it, leading to "less product innovation, fewer consumer choices, and higher prices in the marketplace."⁶⁵ Let's parse those phrases for a moment.

IT IS WORTH EXPLORING
THE LANGUAGE OF THE ORDER
FOR WHAT IT TELLS US
ABOUT THE FTC'S EXPRESS
ANALYTICAL PROCESS.

1. *Pfizer could delay or drop one compound.* This is something that can happen in every merger with a potential overlap in research. If this is the test, it proves too much. No transaction would be allowed. Here, given the failure rate of earlier phase compounds and the fact that

two other companies were much farther advanced in the process, it would make no business sense to drop or delay anything. One should not overlook the distinction between what is theoretically possible, and what a party in the real world is likely to do.

2. *If Pfizer dropped one compound, it would lead to fewer consumer choices.* This is a very odd way to describe competition in medical research. The question is what is likely to work best, on which tumors, with which side effects. Two different compounds are very unlikely to act in identical ways. Arguments about consumer choice assume that cancer therapy is like flavors of chewing gum. And it assumes that each research product will lead to an actual product. Again, the theoretical language is broad, but it doesn't connect to the facts on the ground.
3. *If Pfizer dropped one compound it would lead to higher prices in the marketplace.* There is no supporting data for this astounding characterization of the cancer therapy marketplace. Is the FTC saying that the price of the OSI compound would be lower if the Pfizer compound had come out? And is it saying that the earlier stage compounds would have made it to the market?

As a general rule, first generation products tend to price at parity with each other or close to it (depending upon efficacy, toxicity, and the like). A truly superior product might try to command a premium, but reimbursement these days is so complicated that it is unclear whether even a better product can command a higher price.⁶⁶ When a new generation of products comes along, the older one tends to drop in price. But the factors that constrain pricing on patented prescription drugs in general, and cancer therapies in particular, have nothing to do with the classical economic theories of competition. Often, the major question is not "How many products are out there?" but rather "How much will the government and the insurers pay for a drug that extends life by X months?"⁶⁷ The structure of the prescription drug market, especially as more and more decisions are made by governments and insurers based on cost effectiveness grounds,

means that one has to be very careful about general statements about what “would” lead to higher prices.

None of the stated grounds justified the intervention in this case, and the facts of what later took place in terms of drug approvals confirm that no intervention was needed or useful. The proper approach here would have been a future goods analysis. On that basis, no intervention would have taken place.⁶⁸

6. 2001—*Genzyme/Novazyme*⁶⁹

This is the poster child for pure innovation market analysis. As described in the FTC Press Release:

“Pompe disease is a rare, often fatal, disease affecting infants and children, for which there is currently no effective treatment. Because of the relatively limited number of Pompe patients, therapies for Pompe disease fall under the Orphan Drug Act (ODA). The first Pompe therapy to gain FDA approval will obtain seven years of market exclusivity under the ODA. A second therapy may break that exclusivity only by establishing superiority over the first therapy.”⁷⁰

What is interesting is the debate between Chairman Muris and Commissioner Thompson over the decision to close (in 2004) the investigation of the merger which took place in 2001.

The opening salvo was whether indeed increased concentration leads to decreased innovation. Muris cited work showing that such a link has not been established. This is not surprising, since innovation is not a unitary concept. What encourages innovation in the attempt to find a cure for cancer may well not be the same thing that encourages innovation in the ways to decrease energy use.

Here, only two companies were working in the field. Given that the disease at issue affects a small number of people (i.e. the potential market for any end product is small), and that the research was at the time preliminary, risky, and expensive, it was not likely to draw others to participate in it. And this leads us to the most important part of the Muris opinion—his deep dive into the facts of the case.

At the time of the cases, and the opinions being discussed, there was no treatment for Pompe disease.⁷¹ The issue for Muris was whether the merger was likely to reduce the incentive to invest in the R&D on Pompe disease and whether it was likely to give the merged firm the ability to conduct that R&D more suc-

cessfully.⁷² The question is not, and cannot rationally be, whether gross R&D spending will be reduced. Almost every acquisition or merger does that; it is part of the efficiencies that companies look for when doing a deal. Even where the projects directly overlap, combining them can lead to administrative savings. And this does not even reach the difficult and fascinating question of how to deal with a reallocation of assets—a decrease in R&D for one disease or approach vs. an increase for another. Is this good or bad for innovation, and how would you know? On the facts here, two R&D programs had already failed because they could not produce the enzyme on commercial scale. Genzyme and Novazyme had the remaining two programs.

Genzyme was a significant biotech company, with over 5,000 employees in 2001 and revenues approaching \$1 billion. Novazyme was a relative start up, with no sales and some 80 employees. At the time of the merger, the Novazyme project was in the early pre-clinical stage. Genzyme had tried two joint ventures in the field, and both had failed. As a result, and using the knowledge from those failures, Genzyme was ramping up its own project. At the time of the merger, its compound was also at the early pre-clinical stage.

“SOMETHING WAS DELAYED, THEREFORE YOU INTENDED TO DELAY IT” IS A CLOSE COUSIN TO “ONLY TWO PEOPLE ARE DOING RESEARCH IN THIS FIELD THEREFORE ONLY TWO PEOPLE ARE CAPABLE OF DOING RESEARCH IN THIS FIELD.” THE EXTRAPOLATION FROM OBSERVATION TO CONCLUSION IS UNSUPPORTED.

It bears noting again that for drugs entered in Phase I testing, the failure rate is between about 75-85 percent.⁷³ These compounds were even farther back. It was by no means likely that either of these projects would make it to the finished drug stage.

Muris then looked at the impact of the Orphan Drug Act. In an attempt to encourage companies to research cures and treatments for diseases with small patient populations,

Congress provided a financial carrot. The first drug approved for an Orphan disease gets seven years of market exclusivity. A second drug can break that exclusivity, but only by establishing superiority over the first, a difficult standard.

At the time of the merger one would assume that each company was moving its project as quickly as it could. Post-merger, Genzyme still had the incentive to get a product to market as soon as possible, to start earning a return on its investment. So the question became the nature of the incentive to develop the second product. Genzyme could use the Novazyme compound for a comparative experiment and, allowing for potential synergies, gain the support of the relevant patient advocacy group.⁷⁴

Thompson said that the fact that the Novazyme project had been delayed was evidence that Genzyme intended to delay it.⁷⁵ This kind of odd logic crops up in various contexts.⁷⁶ “Something was delayed, therefore you intended to delay it”

is a close cousin to “Only two people are doing research in this field therefore only two people are capable of doing research in this field.” The extrapolation from observation to conclusion is unsupported. We would hardly say that because a company’s leading project failed, that the company meant for it to do so. Muris disputed Thompson’s reading of the facts.⁷⁷ From a real world perspective, we see several reasons why it would seem irrational for Genzyme to delay development of a second product.

First, anyone who has been involved in pharmaceutical R&D can verify that coming up with firm timelines for clinical trials and FDA action is very difficult. To come up with a timeline for a compound *that is not even in the clinic*, is to engage in wild guesswork. Genzyme would want to have that second product on the market at the latest by the time any ODA exclusivity on the first product expired. There also was evidence that Genzyme wanted to use the technology in the Novazyme program to develop second generation therapy for Pompe disease, and first generation products for other similar disorders.⁷⁸ All of this suggests that there was plenty of motivation to develop the second product as quickly as possible.

Another fact was that the Novazyme’s president was to run the R&D project, and his own son suffered from Pompe disease. His motivation went well beyond economics. Finally, given the length of time of the investigation, there was in effect a two-year look back at actual R&D effort, and no evidence of reduced effort (or spending).⁷⁹ The question for Muris was, one might suppose, the question for the President of Novazyme—which path promised to get an effective treatment for Pompe disease approved and on the market faster—keeping his own project independent, or joining forces with Genzyme.

What Thompson did expressly in his dissent, and others who support the use of broad innovation market analysis have done implicitly, was to assume that an analysis that may have support in one area (i.e. product markets) can be used as if it has support in another area (innovation). They treat innovation as if it was a product market, taking presumptions of anticompetitive effects from the product market realm and applying them to innovation without seeming to acknowledge the difference. Thus, they assume that having two separate research programs is *per se* better than having one, based on the idea that having two widgets on the market is better than having only one.⁸⁰ But as this analysis has tried to show in Part IV, when one tries to test that theory in the real world, it becomes very difficult to explain why more spending or more programs (no matter how weak or ill conceived) are “better” in terms of the anticipated output.

7. 2009 COMMISSIONER ROSCH SPEECH

While it is not a case, the speech by Commissioner Rosch on February 2, 2009 to the ABA Intellectual Property Conference⁸¹ is remarkable both for its candor and for its analysis.

Rosch recognizes that no court has ever invalidated a transaction purely in a purely innovation market (where there was no product at the time).⁸² The FTC raises the issue in cases, but then negotiates settlements. So the question whether an innovation market is cognizable under Section 7 has never been tested. What we have are out-licenses or divestitures of compounds which the parties view as simply a tax on the merger.⁸³

Perhaps the key observation that Rosch makes is that:

“Arguing over whether the parties to a merger have market power in an innovation market is a bit like trying to fit a square peg into a round hole. Traditional market definition analysis is, as a general matter, static by nature....innovation markets are more dynamic...an innovation market cannot be pinned down and it certainly cannot be identified with the certainty the *Philadelphia National Bank* requires.”⁸⁴

Rosch would solve the problem by sliding around it. He would find market power without defining the market first.⁸⁵ On the issue of the two-year window for entry set forth in the Horizontal Merger Guidelines then in effect,⁸⁶ the Guidelines published in 2010 eliminated the problem by eliminating two-year limit entirely⁸⁷ (still leaving the issue of how far out is too far out, of course).

IF PEOPLE BELIEVE THAT THIS
APPROACH WILL LET MERGERS
WITH PALPABLE ANTICOMPETITIVE
RISK GET THROUGH, THEN WE NEED
TO FIND A WAY TO ANALYZE THESE
MERGERS IN A MANNER THAT IS
CONSISTENT, PREDICTABLE,
AND REFLECTS THE REALITY
OF COMPETITION AND
NOT JUST ITS THEORY.

But this creates an interesting counterfactual. Has anyone ever seen a case where the merging parties have argued successfully that despite the fact that they are both in the market with products or have late stage (Phase III) compounds in research, that they should be allowed to merge because there are other companies that have compounds earlier in the pipeline (say Phase I or Phase II)? I have not seen such a case. Those earlier stage compounds are deemed to be too far away, and with too small a chance of success, to be treated as “in the market” for defense purposes. Logically, the same standard should be applied to the intervening agencies.

If people believe that this approach will let mergers with palpable anticompetitive risk get through, then we need to find a way to analyze these mergers in a manner that is consistent, predictable, and reflects the reality of competition and not just its theory.

8. 2010—Pfizer/Wyeth⁸⁸

For the purposes of this article, the interest here is not in the decision itself (no divestiture was required on any human health product or compound), but rather the Statement of the Commission which, in a little over four pages, gave a roadmap of the way the then Commission viewed research based pharmaceutical company deals.⁸⁹ The Commission analyzed the transaction in terms of actual goods, future goods, IP and Innovation.

- **Actual Goods:** In this approach, the Commission recognized that there were a small number of conditions for which Pfizer and Wyeth marketed treatments, but their products were not close substitutes for each other (indeed, the Commission said that the products were not even competitive with each other). Further, an undefined but sufficient number of other companies were competing in same markets, with products that were closer substitutes to the Pfizer/Wyeth products than were the Pfizer or Wyeth products to each other.
- **Future Goods/Future Competition:** This is the section that might have been labeled “innovation market analysis” in an earlier case. The fact that it was treated as a future goods issue is encouraging. The Commission noted that there were a small number of diseases where one company had a product and the other was developing a compound that could compete with that product in the future. The conclusion was that the Pfizer and Wyeth products were unlikely to be sufficiently close competitors to cause problems, and they would compete more closely with products of third parties.⁹⁰
- **Intellectual Property:** Would combining the IP of the two firms create a bar to others working in the fields affected? IP is property and, just like a scarce raw material, access to unblocked research avenues is critical to developing new products. The conclusion was that the bar caused by combining the two companies’ patent portfolios would not cause any greater barrier to entry than the IP held by the parties individually. The merger did not increase the bar.
- **Innovation:** The explanation here is among the fullest that the Commission gave. It first laid out some basic facts about the companies doing research in the pharmaceutical industry:

“Finally, staff evaluated whether the transaction would decrease basic research or the pace of innovation in pharmaceutical markets by eliminating a leader in pharmaceutical research and development; changing the incentives of companies performing pharmaceutical research and development; or reducing the number of potential research, marketing, or funding partners. *Pharmaceutical research and development is a dynamic field with multi-*

ple participants including both large and small traditional pharmaceutical companies, specialty pharmaceutical companies, biotechnology companies, and contract research organizations. The evidence does not indicate that the combination raises antitrust concerns in these respects.”⁹¹

What the FTC said is that there can be no shortcut here; no defining the entire pharmaceutical industry as meeting the standards in the 1995 Intellectual Property Guidelines for defining an innovation market.⁹²

R&D is dynamic, broad based, and worldwide. In every R&D divestiture of which the writer is aware, the program is divested worldwide. And certainly the pool of knowledge and talent is a worldwide one, not simply in terms of hiring employees but in terms of networking people from various companies and universities.⁹³ Individual companies set up networks of collaborations⁹⁴ and broader coalitions have formed.⁹⁵ There are examples of the pooling of data across companies⁹⁶ and sharing information and research with non-profit partnerships targeting one or more diseases.⁹⁷ And this is all in addition to the more traditional partnerships between one company and academic scientists and institutions.⁹⁸

VIII. A View from Brussels

Because the structure of European competition law applicable to agreements and cooperation short of mergers is set up with broad prohibitions but with the possibility for exemptions,⁹⁹ the members of DG Competition also have had to deal with some of the issues highlighted in this article in the process of formulating Block Exemptions (“BEs”) and Guidance documents.¹⁰⁰ The BEs have market-share thresholds, such that if the companies exceed those thresholds the BE does not apply (although that does not mean that the agreement violates the law).

The 1984 and 2000 R&D BEs proposed to have different thresholds apply for the exemption depending on whether the parties collaborating were competing in the relevant market.¹⁰¹ But this raised the primary issue: if the work was in R&D, there was no “market” where you could intelligently measure “market shares.” The Commission went back to what it could measure—the shares of the markets for the existing products that were deemed capable of being improved or replaced by the joint R&D products (if they succeeded).¹⁰² If the R&D was directed at a market where the parties had existing goods capable of being improved or replaced by the R&D project outcome, then the BE would only apply if the combined share of that product market did not exceed 25 percent. If the R&D was directed at a field in which neither party had any products to be replaced or improved, the BE applied regardless of the structure or amount of other competition in that R&D sector.¹⁰³

There is a clear recognition that a market-share type of test cannot be applied directly to R&D (or innovation). In fact, the categories of market share analysis really don't apply to R&D itself, which is why the Commission recognized that if the R&D was aimed at an area where neither company had a product, it could come within the BE regardless of how many other companies were, or were not, working in that field.

When it comes to the more general approach of the Guidelines on Horizontal Cooperation Agreements, both the 2001 and the 2010 versions, there is broader language but still a recognition that any rules have to be tied to something tangible. At Paragraph 114 of the 2010 Guidelines there is the following formulation:

“In the first scenario, which is, for instance, present in the pharmaceutical industry, the process of innovation is structured in such a way that it is possible at an early stage to identify competing R&D poles. Competing R&D poles are R&D efforts directed towards a certain new product or technology, and the substitutes for that R&D, i.e. R&D aimed at developing substitutable products or technology for those developed by the co-operation and having similar timing. In this case, it can be analysed if after the agreement there will be a sufficient number of remaining R&D poles. The starting point of the analysis is the R&D of the parties. Then credible competing R&D poles have to be identified. In order to assess the credibility of competing poles, the following aspects have to be taken into account: the nature, scope and size of possible other R&D efforts, their access to financial and human resources, know how/patents, or other specialised assets as well as their timing and their capability to exploit possible results. An R&D pole is not a credible competitor if it cannot be regarded as a close substitute for the parties' R&D effort from the viewpoint of, for instance, access to resources or timing.”¹⁰⁴

THE COMMISSION RECOGNIZED THAT IF THE R&D WAS AIMED AT AN AREA WHERE NEITHER COMPANY HAD A PRODUCT, IT COULD COME WITHIN THE BE REGARDLESS OF HOW MANY OTHER COMPANIES WERE, OR WERE NOT, WORKING IN THAT FIELD.

The Guidelines also provide examples that are interesting, in large part, because they appear to reflect real world scenarios. For example, Section 142, example 2, deals with a situation where the parties are collaborating on research on a new treatment for a disease, one party has a large share of the existing product market for treatments for that disease, patents are expiring in five years, there are only two other research poles, and yet the deal should and would be cleared.¹⁰⁵ There is recognition in this example and the analysis following it of

the realities of the prescription drug industry that is almost unique in the official literature.

Unfortunately, one looks in vain for any *a priori* way of determining how many R&D poles are enough. While this is no doubt frustrating from a theoretical standpoint, it may indeed reflect the approach to be preferred here. As the example makes clear, the number of R&D poles required to allow clearance of collaboration depends on the facts of the situation—the market, the patent protection, the needs to get the research to fruition. All of these are individualized concerns. What is a “sufficient” number of competing R&D poles will depend on the facts of the case.¹⁰⁶

In 1994 Pfizer signed a joint venture agreement to co-promote Eisai’s product, Aricept (treatment for Alzheimer’s Disease). Both companies had R&D projects in the field, as did seven other companies, at least two of which were on the same time line as the Pfizer and Eisai projects (Eisai being a year or two ahead of Pfizer). The Pfizer compound was assigned to Eisai and kept as a backup, if needed. By the time that the notification was filed to the commission (1998), the Eisai product had already been launched by the parties. The product was the first effective treatment for Alzheimer’s disease, and to the extent that this was a market, Aricept certainly had a dominant share. Out of the seven other companies, only one of the projects led to a successful product shortly after Aricept. The Commission cleared the transaction under then Article 81(3) with a comfort letter.¹⁰⁷

What makes this case unusual is that at the time of the notification, the product was already on the market and succeeding. The Commission, correctly, went back and looked at the agreement at the time that it was made (an *ex ante* approach) and held that while the co-promotion agreement did reduce the number of R&D poles, at the time that the deal was done there were sufficient other poles and, in looking at the potential for exemption, the Commission saw the obvious consumer benefit that the co-promotion arrangement had made in getting the product to market. They judged that the parties should not be penalized for their success in being the first ones to market with an important new therapy.¹⁰⁸

Finally, the 2004 Horizontal Merger Guidelines do not speak of innovation markets as such. They do speak of what appears to be a future goods market, but in terms of changes to a “specific product market” that can be “reasonably predicted.”¹⁰⁹ They go on to state:

“In markets where innovation is an important competitive force, a merger may increase the firms’ ability and incentive to bring new innovations to the market and, thereby, the competitive pressure on rivals to innovate in that

market. Alternatively, effective competition may be significantly impeded by a merger between two important innovators, for instance between two companies with “pipeline” products related to a specific product market. Similarly, a firm with a relatively small market share may nevertheless be an important competitive force if it has promising pipeline products.”¹¹⁰

The key for the analysis here is that the DG Competition approach does tie back to the real, tangible world, which is where effects will have to be measured.

IX. A Proposed Theory of Innovation Markets

Any theory of innovation markets should meet two tests. First, it needs to fit within a broader theory of markets, since it must be consistent with them to avoid an ad hoc, unprincipled approach to its application. Second, the theory needs to deal with the reality of the markets to which it supposed to apply, not just the theoretical constructs about them.

Having reviewed the swings and variations in the application of innovation market theory, and the times when that theory is based on assumptions that simply do not hold in the real world, our analysis drives to a somewhat surprising and modest conclusion. The drafters of the 1995 Intellectual Property Guidelines had it pretty much right. And the FTC in its discussion of the Pfizer/Wyeth merger of 2009 seemed to agree.¹¹¹

But while the traditional FTC application of innovation market theory may be incoherent and frustrating, it does let the agency try to catch matters that do not fit well, or at all, within more traditional categories. It is an ultimate gap filler. This type of thinking flows through many of the FTC Consent Orders discussed earlier where standard antitrust verbiage is used in situations where it really doesn’t apply.¹¹²

A gap filler is not necessarily invalid or illegitimate. But if it truly is to be gap filler, rather than something that will expand without limits to fill any desired enforcement role, there have to be some boundaries on where the theory can go. At the end of the exercise, it should be possible to create a working taxonomy or classification system that will enable us to see when innovation market analysis is appropriate, and how to do that analysis.

AT THE END OF THE EXERCISE,
IT SHOULD BE POSSIBLE TO
CREATE A WORKING TAXONOMY
OR CLASSIFICATION SYSTEM THAT
WILL ENABLE US TO SEE WHEN
INNOVATION MARKET ANALYSIS
IS APPROPRIATE, AND HOW
TO DO THAT ANALYSIS.

A. ACTUAL GOODS

This is a standard antitrust analysis. When there are existing goods, the agencies can base a case on them using established and tested principles. There is no need to go searching for other theories to use.¹¹³

B. FUTURE GOODS

Where there is a product on the market and a future product in research, is entry of the latter sufficiently certain and timely to make it part of the product market for analytical purposes? In terms of timeliness, the 1992 Horizontal Merger Guidelines set a two-year limit on entry to be considered part of the market, and the 2010 replacements, while eliminating the two years, keep the concept of timely entry as defining a market participant.

In terms of certainty, at least as far as the pharmaceutical industry is concerned (and recall that is where almost all of the innovation market cases take place), compounds at Phase III and above would seem to be a rational cut-off point (greater than 50 percent chance of success; time to approval 2-4 years).¹¹⁴ There might be some flex in the definition, depending on the facts of a given situation. If the FDA is reviewing and approving drugs faster for a given disease or unmet need, then there may be a good reason for including Phase II compounds as future goods.¹¹⁵

Future goods is an underutilized category, often improperly slighted in favor of innovation market analysis. All of the groundwork for such an analysis was present in *Roche/Genentech* and *AHP/Cyanamid*. Had the agency applied a future goods analysis, it would have concluded that no intervention was required and, indeed, the potential products were so far away from the market that the risk of both of them even coming to market was so remote that requiring a remedy was unjustified and simply added to the risk that no product would survive.

And remember our earlier counterfactual. Logically, either a compound is close enough to the market to “count” or it is not, regardless of whether the view is from the FTC or the merging parties. I have seen no case where the merging parties have argued successfully that, despite the fact that they are both in the market with products or have late stage (Phase III) compounds in research, they should be allowed to merge because there are other companies that have compounds earlier in the pipeline (say Phase I or Phase II)? Those earlier stage compounds are deemed to be too far away, and with too small a chance of success, to be treated as “in the market” for defense purposes. For the same reason, those earlier stage compounds should not “count” to justify agency intervention.

C. INTELLECTUAL PROPERTY

An argument that a merger creates a patent blockade greater than the patent estates of the individual participants is not always a simple one to prove. But

assuming that the factual hurdle can be jumped, there is no theoretical reason to treat IP as different from any other kind of property.¹¹⁶ But if the agencies are talking about IP, there is no need to talk about innovation markets. Patents are things that one can count, read, buy, sell, and license. It may not be easy to monopolize an IP market. But one does not make the analysis any easier or any better by dragging in innovation.

That leaves the last category, the last block in the square. We are left with innovation, and how to deal with it.

D. INNOVATION/RESEARCH & DEVELOPMENT

It is highly unlikely that pure innovation represents a market that would be defensible on traditional competition law terms, much less one in which one could calculate market shares and Herfindahl indices. Perhaps this is one reason why no pure innovation market case has ever reached a court decision.¹¹⁷ Indeed, there are major problems even trying to define what is meant by innovation, how it could or should be measured, and how much innovation is better or worse than any other amount. Even then, there is the question of how much innovation is out there, or available, and that includes the internet's existence, the linked-in scientific community, and the ability of any company with money to access the relevant science.

BUT IF THE AGENCIES ARE
TALKING ABOUT IP, THERE IS
NO NEED TO TALK ABOUT
INNOVATION MARKETS.

This may well be why the IP Guidelines, and the European Commission Block Exemptions, came at the issue from the flank. They limited innovation market inquiries to cases where the parties have unique access to necessary tangible assets; where the capability to engage in the relevant R&D can be associated with specialized assets or characteristics of specific firms. This is the key. Once the analysis gets back to looking at tangible assets, one can ask what is required to do the research, who has access to such assets, and whether others can get such access. The analysis is back on solid ground.

X. The Revised Innovation Market Theory Applied

This analysis leads to a theory that is both internally consistent and consistent with the external reality of the marketplace: an innovation market analysis is only applicable when the facts do not permit analysis in terms of actual goods, future goods, or IP, and then only applies where there is limited access to necessary tangible assets in order to work in the field.

Would the application of the proposed new theory have made a difference in the case outcomes and, if so, how and why? Hindsight provides an enormous

advantage in making this analysis. It allows a look at what actually occurred in the marketplace—to see whether the remedy applied did, in fact, lead to increased competition, more products on the markets, and all of the attendant benefits that innovation market intervention is supposed to provide.

Based upon that review, of the eight key cases that were reviewed above, the results would not have changed in five of them (*Smog Control Devices*; *U.S. v. G.M.* (truck transmission); *Glaxo/Wellcome*; *Ciba/Sandoz*; and *Genzyme/Novazyme*), although the rationale for intervention or non-intervention would have been different in some.

In the three cases where the result would have changed (*Roche/Genentech*; *American Home/American Cyanamid*; and *Pfizer/Warner Lambert*) the approach presented here counseled against the intervention that took place. Had the FTC looked at the cases as future goods matters, they would have recognized that no intervention was justified. And in each case the factual look back supports such a non-interventional approach.

At the end of the day, the question is whether competition law agencies should intervene in R&D at a very early stage based on what is almost a theological belief that society is better off with two small projects than one larger one. The underpinnings of that belief are shaky, even if there was an agreed upon measuring rod for R&D, apart from looking at what products actually make it to market.

WHAT HAS TO BE REALIZED AND ACKNOWLEDGED IS THAT THERE IS A VIRTUALLY INFINITE SET OF MEDICAL PROBLEMS TO BE RESEARCHED. THE AREAS THAT HAVE LARGER POTENTIAL PATIENT POPULATIONS AND POTENTIAL FINANCIAL RETURN WILL ATTRACT GREATER R&D EFFORTS.

For example, where the scientific problems are extremely difficult, even large companies have found it more productive to pool their resources rather than exploring every dead end alone. Perhaps the most famous example of this is the 1993 Inter-Company Collaboration for AIDS Drug Development.¹¹⁸ And in 2010, companies agreed to share data on clinical trials in Alzheimer’s drug testing.¹¹⁹

This is not to suggest that it would be good policy to force the creation of one large pharmaceutical company. But it is to say that we should be wary of intervening in the decision of these companies to allocate their capital and their efforts in one area rather than another. It should not shock us that very few firms choose to invest in research to find a cure for Pompe Disease. What has to be realized and acknowledged is that there is a virtually infinite set of medical problems to be researched. The areas that have larger potential patient populations and potential financial return will attract greater R&D efforts.

The narrower one defines the market, the fewer players one will have. Thus, a field defined as “R&D into blocking cancerous tumor growth,” will have many

participants. A field defined as “Impeding drug cell signaling pathways impacted in rapid cell division” (a subset of the first field), will have fewer participants. This is simply a function of how analysis works—no more and no less. It is one approach out of many. There is nothing malignant or even mysterious about this. The narrower the focus, the fewer the objects there will be in the field. If the question posed was how many companies were working on a cure for Pompe Disease using compound NZ-1001, there was only one member of that set; Novazyme. But that fact tells us very little by itself.

XI. A Proposal

Innovation market theory, as it has been applied to date, rests on a flawed foundation. It is a conceptual stretch to cover the situation where more established theories do not seem to apply. And, at least as far as the research-based pharmaceutical industry is concerned, the theory relies on assumptions about how these companies behave that are contradicted by the facts that drive behavior in the marketplace.

In its analysis of cases to date, the FTC seemed to be unduly concerned that transactions might eliminate competition between two or more early stage development projects even when history demonstrated it was highly unlikely that either (much less both) project(s) resulted in a product on the market. Recognizing that a traditional future goods analysis did not support intervention (and therefore did not solve the perceived problem), the agency stretched the future goods rules by cloaking them in innovation market language. But, rather than increasing innovation, that approach may well have hindered it. Once it is recognized that there is no necessary harm in these cases, the need to stretch to find a remedy goes away.

The approach suggested here is one of humility and practicality. There is a role for innovation market analysis, but it is a modest one. Rather than constitute a free-roving charter to substitute the judgment of antitrust regulators for decisions of the private parties involved, it should be used to allow intervention where such action can be justified in terms of practical tangible impact.

Economists and lawyers have experience with traditional actual goods markets. There is a large body of data on prices, demand, and firm behavior. There are data on future goods and the impact of goods on the edge of the market as well as on the behavior of participants with goods on the market (and real time frames associated with that data—which is what led to the two-year clause in the

RECOGNIZING THAT A
TRADITIONAL FUTURE GOODS
ANALYSIS DID NOT SUPPORT
INTERVENTION (AND THEREFORE
DID NOT SOLVE THE PERCEIVED
PROBLEM), THE AGENCY
STRETCHED THE FUTURE GOODS
RULES BY CLOAKING THEM IN
INNOVATION MARKET LANGUAGE.
BUT, RATHER THAN INCREASING
INNOVATION, THAT APPROACH
MAY WELL HAVE HINDERED IT.

1992 Merger Guidelines). And once it is accepted that IP is a form of asset with certain definable characteristics, antitrust lawyers and economists can talk about how to avoid multiplication of the statutory grants through merger. But as Commissioner Rosch noted,¹²⁰ innovation is a very different kind of animal.

Unless an analysis ties innovation to output, there is no verifiable way to know what to measure, how to measure it, how to encourage it, or what the optimum conditions are for it to grow and flourish. If one university hires five experts on the causes of Alzheimer's disease, does that speed up, retard, or leave unchanged the time line for coming up with an effective cure? What is the basis for your answer? If an observer hopes that the mass hiring speeds up the finding the cure process, would he or she say the same thing if one company hired those same five scientists? What if one company partnered with five universities? What if five companies pooled their resources?

Asking these questions throws a light on an underlying core issue. The question isn't so much whether one deal is good or bad, or even whether it helps innovation or retards it. The question is how one would ever be able to predict the outcome with any degree of confidence. The FTC has jumped that question by making presumptions about how the parties would or should behave. But those presumptions have been shown to be unsupported, leaving the issue of showing a potential benefit from intervention open.¹²¹

So the conclusion of this analysis and the look back at applicable cases is a plea for a bit of humility on the part of the competition law enforcement groups. Where there are actual goods markets, future goods markets (properly defined) or IP markets, then the agencies can apply their traditional theories and have some confidence in the outcome. But when one looks at innovation and the innovative process, it is crucial to recognize that there is much that simply is not known. On the taxonomy and innovation market definition suggested here, the analysis ties to limited physical assets. Those can be found, counted, and costed out. But to go further, and to continue to try to control the actual innovative process itself by applying theories and presumptions, risks doing far more harm than good. ▼

-
- 1 Speech by Commissioner Rosch on February 2, 2009 to the ABA Intellectual Property Conference; <http://www.ftc.gov/speeches/rosch/090205innovationspeech.pdf> (hereinafter "Rosch Speech") at 13-14. This may be why the theory has never been asserted successfully in a litigated case.
 - 2 It is not always simple to answer these questions. But that is not an excuse for a failure to try. Indeed, the attempt itself gives us some valuable information about the theories in play.
 - 3 1995 Antitrust Guidelines for the Licensing of Intellectual Property ("IP Guidelines") §3.2.3, *available at* <http://www.justice.gov/atr/public/guidelines/0558.htm>
 - 4 The literature on innovation market analysis is rich and full. While we will spend most of our time examining the cases themselves, we will make multiple citations to certain works: Abrantes-Metz et

al, *Empirical Facts and Innovation Markets: Analysis of the Pharmaceutical Industry*, ANTITRUST SOURCE 1 (March 2005) (hereinafter "Abrantes-Metz"); Carrier, *Two Puzzles Resolved: Of the Schumpeter-Arrow Stalemate and Pharmaceutical Innovation Markets*, 93 IOWA L.R. 393 (2008) (hereinafter "Carrier"); Gotts & Rapp, *Antitrust Treatment of Future Goods*, ANTITRUST 100 (Fall 2004) (hereinafter "Gotts and Rapp"); and Rapp, *The Misapplication of the Innovation Market Approach to Merger Analysis*, 64 ABA ANTITRUST L.J. 19 (1995) (hereinafter, "Rapp"). For those interested in perhaps the fairest and certainly the deepest analysis done in the field, see MARCUS GLADER, *INNOVATION MARKETS AND COMPETITION ANALYSIS: EU COMPETITION LAW AND US ANTITRUST LAW* (2006) (hereinafter "Glader").

- 5 U.S. v. Automobile Manufacturers Association, 307 F. Supp. 617 (C.D. Cal 1969), appeal dismissed sub nom City of New York v. U.S., 397 U.S. 248 (1970).
- 6 U.S. v. General Motors Corp., Civ. No.93-530 (D. Del. filed 11/16/93) (hereinafter "GM Complaint").
- 7 See speech of Anne. K. Bingaman, Antitrust and Innovation in a High Technology Society (January 10, 1994); www.usdoj.gov/atr/public/speeches/0108.htm.
- 8 Back in the 1960s and 1970s the Department of Justice made a similar attempt to create law out of speeches about patent licensing terms. The approach was to convince companies not to use certain terms in patent licensing by simply stating that such terms were illegal, without actually having to bring and win any cases. It was referred to as "Luncheon Law," as the speeches often followed a lunch. Neither then, nor now, did it provide anything in the way of rigorous analysis. See Bernard, *The 2008 EC Sector Inquiry Regarding Pharmaceuticals: What Does It Mean From a Research-Based Company Perspective*, GLOBAL COMPETITION POL'Y 10-11 (November 2008) at pages 10-11. Richard Rapp believed that in most cases invoking "innovation markets" was just a way of talking about future products/potential competition but going farther back into the R&D pipeline; Rapp, *supra* note 4 at 2. Indeed, the cases are consistent with such a definition. But Rapp did not mean that such an approach was valid or correct. Those are issues that will be explored in the course of this article.
- 9 IP Guidelines, *supra* note 3, §3.2. It is interesting that the Guidelines, at §3.1, phrase their concern in terms of:

An arrangement that effectively merges the research and development activities of two of only a few entities that could plausibly engage in research and development in the relevant field.... (emphasis supplied).

The focus is on companies that *could* do work in the field, not simply those that happen to be working there at a given point in time. This point is critical, and has all too often been overlooked or ignored in the cases.
- 10 One way to look at this is to consider innovation as a driver of economic growth, as many have. See Bernard & Tom, *Antitrust Treatment of Pharmaceutical Patent Settlements*, 15 FEDERAL CIRCUIT BAR J. 617, 618 (2006); Carrier, *supra* note 4 at 399 and note 8. But this is still an abstraction. What is being acquired or divested is something specific, and should lead to something concrete at the end of the day.
- 11 Rosch speech, *supra* note 1 at 9.
- 12 Perhaps the most notable example of this point is far outside of the competition law universe, i.e. the Manhattan Project to develop the atomic bomb in World War II. Multiple projects were yoked together and coordinated by the government with the end of developing a workable bomb as soon as possible. See <http://www.cfo.doe.gov/me70/manhattan/> ; <http://nuclearweaponarchive.org/Usa/Med/Med.html>. A single project was deemed to be the most efficient and the best way to get to the goal of having a workable "product" for the market. But note that once the debate shifts to how to best get a research project to market, we are talking the language of future goods markets, not innovation *per se*.

- 13 This is a point that comes back strongly in Commissioner Muris' opinion in the Genzyme case *available* at <http://www.ftc.gov/os/2004/01/murisgenzymestmt.pdf> (hereinafter "Muris Opinion") at 2-3, 5-6. From an intervention standpoint, the approach may best be asymmetric—if there are many people working in a field, then the presumption should be to let parties determine their own allocation of research capital and time. But the converse does not mean that action should be taken. The fact that there are relatively few people actually working in a field is not a sufficient cause for intervention.
- 14 Some analyses seem to want to do it in reverse—enough R&D is that amount that provides for the (eventual) launch of more than one product in a field. The problem with this is that it is not applicable *ex ante*. At the time that decisions are being made about requiring divestitures the theory does not provide us with any way to predict whether such divestitures will be helpful or harmful.
- 15 *But see* Rapp, *supra* note 4 at 34 and the commentary on the amount of money that GM spent over time and the lack of reward. This may be another asymmetrical situation—if you spend little money, you may not get results. But simply spending a lot of money doesn't guarantee any better outcome.
- 16 How do you distinguish a major invention from a minor one? How do you balance them?
- 17 If you adopt this approach, you then need to figure out how to compare cell phones, cameras, deep seas drilling tools, and prescription drugs.
- 18 Also, Goodhart's Law cautions us that once a social or economic indicator or other surrogate measure is made a target for the purpose of conducting social or economic policy, it then will lose the information content that would qualify it to play such a role. See http://lesswrong.com/lw/1ws/the_importance_of_goodharts_law/. While originally applied to monetary policy, it has broader meaning.

That is, once you start measuring GDP as a way of gauging social welfare, people will start to figure out ways to make GDP go up without improving social welfare (say, by swapping dirty financial derivatives). Once Google starts measuring inbound links as a way of evaluating the importance of web-pages, people will figure out how to increase the inbound links to unimportant pages (splogging, blogspam). And once you measure fat or calorie content as a proxy for the healthfulness of food, manufacturers will figure out how to decrease fat and calories without making the food more healthful (reducing fat by adding sugar, reducing calories by adding poisonous artificial sweeteners). <http://boingboing.net/2010/04/29/goodharts-law-once-y.html>.

In the current case, if the number of compounds in development is "the" measure of innovation, then Goodhart's Law teaches that we can expect that more compounds will be generated. What it will not say is whether that greater number of compounds truly correlates with greater innovation, other than in the tautological sense that "higher number equals more innovation" by definition.
- 19 This is something that our European colleagues seem to have accepted. See 2010 Draft Horizontal Cooperation Guidelines, *infra* note 104, at §§10, 41, and 106. The 2010 revised version of the DOJ/FTC Horizontal Merger Guidelines, *infra* note 87, tries to suggest ways to minimize the importance of the market. While a full discussion of this debate is beyond the scope of this article, it is important to note that the statute speaks of a "line of commerce." Softening guidelines doesn't change the underlying law.
- 20 Since there are no existing products in the innovation market analyses that have been put forth, economic hypotheses based on pricing impacts and diversion ratios logically have no application here.
- 21 See Rapp, *supra* note 4 at 36.
- 22 IP Guidelines, *supra* note 3, §3.2.3.
- 23 Rapp, *supra* note 4 at 37.

- 24 In the author's experience, when Pfizer acquired Pharmacia in 2003, Pfizer fought very hard to retain an agreement that Pharmacia had with Altana to develop roflumilast, viewing the compound as a potential complement in the treatment of Chronic Obstructive Pulmonary Disease. While Pfizer succeeded at the FTC, the compound then did not succeed in the clinic. In 2005 Pfizer terminated the agreement and the compound reverted to Altana; see *Daxas deal leaves Altana short of breath, available at* http://www.pharmiweb.com/features/feature.asp?ROW_ID=624 .
- 25 While someone might suggest that it is theoretically possible to sign up all of the key researchers in a field to long-term exclusive employment contracts, given the breadth of science around the world, this risk does not seem to be a realistic possibility.
- 26 Carrier, *supra* note 4. While the conclusions reached by Carrier are not the same as those reached in the current article, I adopted his approach of deriving a theory and then testing it against what actually happened in the real world.
- 27 *Id.* at 418-420; see also Gots & Rapp, *supra* note 4 at 101. If there is anything like a consensus in the field, this is it.
- 28 See, e.g., Case COMP/M.1846, Glaxo Wellcome/SmithKline Beecham (2000) at ¶190, *available at* <http://lib.hebust.edu.cn/ywyfzsk/zsk/pharm-docum/b011.pdf> .
- 29 See sources collected in Carrier, *supra* note 4 at 411-414.
- 30 See, e.g., Grabowski, *Patents, Innovation and Access to New Pharmaceuticals*, Duke University, 2002; *available at* http://levine.sscnet.ucla.edu/archive/grabow-patents_innov.pdf.
- 31 See, e.g., DiMassi et al, *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151,162 (2003).
- 32 There are journals dedicated to following these developments and deals. See, e.g., FIERCE BIOTECH, a daily on-line publication *available at* http://www.fiercebiotech.com/?utm_medium=nl&utm_source=internal. The flow of alliances and acquisitions is unending. The facts simply do not support any assumption that a few large companies are the only ones capable of or doing research in a field.
- 33 See GW Law School report at <http://www.law.gwu.edu/Academics/FocusAreas/IP/Pages/Cloning.aspx>.
- 34 Carrier, *supra* note 4 at 401 (citing the IP Guidelines Section 3.2.3).
- 35 For example, Pfizer paid \$1.3 billion to acquire a company, Esperion, that had one promising phase II compound <http://www.cnn.com/2003/BUSINESS/12/21/us.pfizer.reut/>. This is not unique to Pfizer. See generally <http://www.businesschemistry.org/article/?article=113>. And the saga of the acquisition of OSI is instructive. See text accompanying notes 62 - 64, *infra*.
- 36 Price competition among generic drug sellers in the United States is vicious. Many large chains in the United States are now offering a 30 day supply of the most popular generic drugs for well under \$10, and some at half that price or less. See the report of the National Conference of State Legislatures (January 2009), *available at* <http://www.ncsl.org/IssuesResearch/Health/GenericDrugPricingandStates2009edition/tabid/14440/Default.aspx>. See also <https://webapp.walgreens.com/MYWCARDWeb/pdf/Value-PricedGenericsList.pdf>.
- 37 See, e.g., the impact on a major patented cholesterol lowering product when a competitor went off patent; http://www.forbes.com/2007/07/18/pharmaceuticals-pfizer-lipitor-biz-sci-cx_mh_0718_pfizer.html
- 38 This issue is not unique to any one company. It is simply a fact of life in the research-based drug

industry that even late stage projects fail. Pfizer had a promising compound in phase III for boosting “good” cholesterol. It would have opened up a new market, and complimented an existing product that lowered “bad” cholesterol. On November 30, 2006 the Pfizer CEO declared that the compound would be a potential blockbuster. On December 2, 2006 the project was killed based on side effects that had just come to light. The project cost over \$800 million by the time it ended. <http://www.fiercepharma.com/special-reports/pharmas-biggest-flops/torcetrapib-pharmas-biggest-flops>. See also Eli Lilly and Bristol-Myers Squibb regarding the failure of Erbitux for treatment of colon cancer; http://www.dnaindia.com/health/report_colon-cancer-drug-failure-challenges-assumptions_1392852, and Novartis and Antisoma with respect to their Phase III compound for non-small cell lung cancer, http://www.dnaindia.com/health/report_colon-cancer-drug-failure-challenges-assumptions_1392852.

- 39 See Novartis, and the successor to its very successful product, Gleevec, http://www.fiercepharma.com/story/data-aids-novartis-push-replace-gleevec-tasigna/2010-06-04?utm_medium=nl&utm_source=internal.
- 40 Genzyme is a case study on real world facts. The issue was treatment for Pompe Disease, a fatal condition affecting a relatively small number of people, for which there was no treatment. The only companies working in the field were Genzyme, which had experience with the type of approach involved, and Novazyme. For reasons that will be discussed later, the FTC elected not to challenge the deal. On the facts, that seems to have been the right decision (although one Commissioner dissented on what we can call traditional antitrust grounds about not allowing mergers to monopoly). This case, and the theoretical battle over what is best for innovation, will be discussed further below.
- 41 Abrantes-Metz, *supra* note 4 at 5; Rapp & Gotts, *supra* note 4 at 101.
- 42 *Guidelines on the Application of Article 101 of the Treaty on the Functioning of the European Union to Horizontal Cooperation Agreements*, at ¶¶117-120; available at http://ec.europa.eu/competition/consultations/2010_horizontals/guidelines_en.pdf. See also the case law cited earlier, *supra* note 28.
- 43 IP Guidelines *supra* note 3, Section 2.1. There may be interesting factual issues where a patent covers more than one area, or has application in more than one area. But this can arise with any asset.
- 44 It is encouraging that this analysis is both consistent with, and helps to explain, the FTC clearance and analysis in the recent Pfizer/Wyeth transaction; see Statement of Federal Trade Commission Concerning Pfizer/Wyeth, No.091-0053, available at <http://www.ftc.gov/os/caselist/0910053/091014pwyethstmt.pdf>.
- 45 There is a rich library of work on the distinction between Type 1 errors (prohibiting something that should be allowed) and Type 2 errors (allowing something that should be prohibited). See *generally* INTERNATIONAL ANTITRUST LAW AND POLICY, THE 2008 FORDHAM COMPETITION LAW INSTITUTE (B. Hawk, ed.) at Chapters 16-19 (articles by John Fingleton & Ali Nikpay; David Lewis; Paul Lugard; and Daniel Rubinfeld). That debate is beyond the scope of this article. For our purposes, the key is that given the inability to define the conditions for encouraging innovation, intervention should be a last, rather than first, resort.
- 46 IP Guidelines, *supra* note 3, Section 3.2.3; Carrier, *supra* note 4 at 401.
- 47 For an example of what is out there, see WORKING TOGETHER, CREATING KNOWLEDGE (The University-Industry Research Collaboration Initiative)(2001) available at <http://www.acenet.edu/bookstore/pdf/working-together.pdf>.
- 48 The reference is to a legendary children’s book (and movie), *The Wizard of Oz*, about a little girl named Dorothy who is whisked away from her home in Kansas by a tornado and deposited in the magical Land of Oz. After numerous adventures, she makes it home to Kansas safe and sound. See *generally* <http://www.imdb.com/title/tt0032138/synopsis>

- 49 In re: Roche Holding Ltd., 113 FTC 1086 (1990).
- 50 Carrier, *supra* note 4 at 430-431.
- 51 See Abrantes-Metz, *supra* note 4 at 5; Carrier, *supra* note 4 at 416-419.
- 52 Carrier, *supra* note 4 at 431.
- 53 In re: American Home Products, 119 FTC 217 (1995).
- 54 Carrier, *supra* note 4 at 432.
- 55 *Id.* at 432-433.
- 56 In re: Glaxo, 119 FTC 815 (1995).
- 57 See <http://www.centerwatch.com/drug-information/fda-approvals/drug-details.aspx?DrugID=348>.
- 58 By 1999 Glaxo still had 83 percent of the sales of migraine treatment products, with Imitrex. Zomig (the former Wellcome compound) did reach the market, and had a 7 percent share at this time. By 2006 Imitrex still had a 56 percent share, but three other companies had share of at least 10 percent each. See Carrier, *supra* note 4 at 434.
- 59 In re: Ciba-Geigy Ltd., 123 FTC 842 (1997).
- 60 In re: Pfizer Inc., Case No. 001-0059 (2000), *available at* <http://www.ftc.gov/os/caselist/c3957.shtm>. Note that when the same acquisition was considered by the European Commission, the authorities there did not agree with the FTC's view either on market definition or on the impact of phase I compounds. The EC found no remedy required in the oncology field. See Case No COMP/M.1878, Pfizer/Warner Lambert at ¶¶42, 77-80 (2000) *available at* http://ec.europa.eu/competition/mergers/cases/decisions/m1878_en.pdf.
- 61 <http://www.centerwatch.com/drug-information/fda-approvals/drug-details.aspx?DrugID=867>.
- 62 <http://www.thepharmaletter.com/file/47058/roche-and-genentech-sign-oncology-deal-with-osi-valued-at-187-million.html>.
- 63 <http://www.businessweek.com/ap/financialnews/D9G7QS901.htm>.
- 64 <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004/ucm108383.htm>.
- 65 FTC Complaint, ¶29.d. at <http://www.ftc.gov/os/2000/06/pfizercmp.htm>.
- 66 See, e.g., http://www.fiercepharma.com/story/gsk-cancer-pill-tyverb-falls-short-nice/2010-06-09?utm_medium=nl&utm_source=internal.
- 67 See, e.g., the decision of the U.K agency denying reimbursement on cost effectiveness grounds of several new compounds; http://www.fiercepharma.com/story/avastin-gets-new-thumbs-down-nice/2010-07-09?utm_medium=nl&utm_source=internal; <http://www.medscape.com/viewarticle/712971>. See also the debate of U.S. healthcare officials deciding on whether to pay for a prostate cancer vaccine; http://www.fiercevaccines.com/story/dendreon-shares-routed-news-medicare-provenge-review/2010-07-01?utm_medium=nl&utm_source=internal.

- 68 Interestingly, the European Commission took a position very much like the one recommended here in the Pfizer/Warner Lambert case, holding that the competition in later phase compounds made any adverse impact from the acquisition too speculative to base action upon. See Case No COMP/M.1878, Pfizer/Warner Lambert at ¶¶42, 77-80 (2000) available at http://ec.europa.eu/competition/mergers/cases/decisions/m1878_en.pdf.
- 69 FTC Closes its Investigation of Genzyme Corporation's 2001 Acquisition of Novazyme Pharmaceuticals, Inc. (2004); <http://www.ftc.gov/opa/2004/01/genzyme.shtm>.
- 70 *Id.*
- 71 In April 2006 the FDA approved the first such treatment, Genzyme's Myozyme; <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108645.htm>.
- 72 Muris Opinion, *supra* note 13 at 6.
- 73 *Id.* at note 29.
- 74 *Id.* notes 42 and 43.
- 75 Thompson Statement In Genzyme, at 5, available at http://www.ftc.gov/os/2004/01/thompson_genzymestmt.pdf.
- 76 Logicians refer to it as the fallacy of "post hoc, ergo propter hoc" which loosely translates as "After this, therefore because of this." See <http://www.skepdic.com/posthoc.html>.
- 77 Muris Opinion, *supra* note 13 at 23.
- 78 Genzyme Press Release of August 7, 2001, quoted at Muris Statement, *supra* note 13, at note 34.
- 79 Muris Opinion, *supra* note 13 at 17.
- 80 This is not unique to Thompson. It flows through many of the supporters of broad use of innovation markets. See, e.g., Dennis A. Yao & Susan S. DeSanti, *Innovation Issues Under the 1992 Merger Guidelines*, 61 ANTITRUST L.J. 505, 510 (1993) (speaking of preserving innovation is the benefits that arise from having multiple paths towards the same goal). And see *In re: American Home Products*, 119 FTC 217 (1995) at ¶13 (speaking of the HHI for a market for the development of something that does not yet exist); available at http://www.ftc.gov/os/decisions/docs/vol119/FTC_VOLUME_DECISION_119_%28JANUARY_-_JUNE_1995%29PAGES_217-315.pdf.
- 81 Rosch speech, *supra* note 1.
- 82 *Id.* at 13-14 and note 37.
- 83 One primary reason that companies agree to such a tax is, again, a factor of real world business. They want to close their deal. The longer things stand open, the more disruption there is, the more good people they lose, and the longer it takes to get the businesses integrated. See, e.g. <http://seeking-alpha.com/article/162831-delay-in-oracle-sun-merger-hurts-both-parties>; <http://www.v3.co.uk/v3/news/2251658/sun-cuts-jobs-blaming-merger>. Further, a major transaction may not hold in place for the six months or more that it may take to get an FTC decision, and closing simply on the basis of no Federal Court injunction stopping you is a risky process. Finally, once the parties agree to the terms of a settlement, they have little interest in delaying things further by arguing over the terms of and theories underlying the proposed complaint. See, Kovacic & Winerman, *Competition Policy and the Application of Section 5 of the Federal Trade Commission Act*, 76 ANTITRUST L.J. 929, 941 note 36 (2010).

- 84 Rosch speech, *supra* note 1 at 21.
- 85 *Id.* at 22.
- 86 1992/1997 Horizontal Merger Guidelines Section 3.2, available at <http://www.justice.gov/atr/public/guidelines/hmg.htm>.
- 87 See 2010 Horizontal Merger Guidelines, available at <http://www.ftc.gov/os/2010/08/100819hmg.pdf>.
- 88 <http://www.ftc.gov/os/caselist/0910053/index.shtm>.
- 89 <http://www.ftc.gov/os/caselist/0910053/091014pwyethstmt.pdf>.
- 90 Admittedly, the conclusions were based on non-public data so it is difficult to evaluate how far the Commission has moved towards a predictable theory of future goods and when a compound should “count” for this purpose. But the frame of the analysis is right.
- 91 FTC Statement in Pfizer/Wyeth, *supra* note 89 at 4, emphasis supplied.
- 92 Intellectual Property Guidelines, *supra* note 3, Section 3.2.3.
- 93 See, e.g., <http://www.biomedexperts.com/> ; <http://www.researchgate.net/>. Indeed, the number of networks and strategic alliances has grown rapidly in recent years. See generally Glader, *supra* note 4 at 35.
- 94 See, e.g., http://www.almirall.com/webcorp2/cda/lmD_05.jsp.
- 95 See The Inter-Company Collaboration for AIDS Drug Development, at <http://www.ncbi.nlm.nih.gov/pubmed/11362548>.
- 96 See, e.g., the sharing of data on attempted treatments for Alzheimer’s disease, at http://www.fiercebiotech.com/story/big-pharma-share-failed-alzheimers-drugs-results/2010-06-11?utm_medium=nl&utm_source=internal.
- 97 See, e.g., malaria at <http://query.nytimes.com/gst/fullpage.html?res=9A01E6D81F3EF932A35750C0A9619C8B63&scp=3&sq=drug%20company%20partnerships&st=Search>; growth of public/private partnerships for diseases of poverty at <http://content.healthaffairs.org/cgi/content/full/24/4/1057>; lymphatic filariasis at <http://www.gsk.com/community/filariasis/index.htm>; and multi-drug resistant tuberculosis at <http://www.lillymdr-tb.com/>.
- 98 The list here is immense. See, e.g., GSK with Harvard Stem Cell Institute, <http://harvardscience.harvard.edu/foundations/articles/glaxosmithkline-and-harvard-stem-cell-institute-announce-major-collaboration-ag> ; and Pfizer with Washington University, http://media.pfizer.com/files/news/press_releases/2010/wu_collaboration_051710.pdf.
- 99 U.S. law, by virtue of the “rule of reason” (only “unreasonable” restraints are barred) did not lend itself to the exemption process. See, e.g., Bernard, *Private Damages Actions: A U.S. Perspective on Importing U.S. Damages Actions to the EU*, eCCP (October 2007).
- 100 A full discussion of EU legal structure is well beyond the scope of this article. The EU process was originally set up whereby almost any joint action by competitors could be viewed as a violation of the conspiracy in restraint of trade provision (now article 101), but the system had a process for requesting an “exemption” based on the net pro-competitive nature of the conduct at issue. The Commission was flooded with requests for exemptions, and came up with the idea of issuing Block

Exemptions (“BEs”) such that if your agreement fit within the four corners of the BE you were safe without having to make any individual request. The entire individual request process was abolished later, but the idea of the BE remains firmly embedded.

- 101 1984 R&D BE located at http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!DocNumber&lg=en&type_doc=Regulation&an_doc=1985&nu_doc=418; 2000 R&D BE located at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:304:0007:0012:EN:PDF>.
- 102 *Id.* at §3.
- 103 *Id.* at §4. There is a full discussion of this whole area in Glader, *supra* note 4 at 75-84, and the Commission lays out a concise analysis in the *2010 Horizontal Cooperation Guidelines* at §3.2; http://ec.europa.eu/competition/consultations/2010_horizontals/guidelines_en.pdf.
- 104 *2010 Draft Guidelines on Horizontal Cooperation*; available at http://ec.europa.eu/competition/consultations/2010_horizontals/guidelines_en.pdf. The language from section 51 of the 2001 Guidelines is the same.
- 105 http://ec.europa.eu/competition/consultations/2010_horizontals/guidelines_en.pdf at ¶141, example 2 (page 39). The definition of an R&D Pole is somewhat opaque. It perhaps can best be analogized to a research project in the broad sense, including the access to financial and human resources, patents, know-how, any other specialized assets, and the capacity to exploit the results. See, e.g., http://www.slaughterandmay.com/media/64578/the_eu_competition_rules_on_horizontal_agreements_apr_2010.pdf at §2.6. Another analogy would be the 1995 (U.S.) IP Guidelines, *supra* note 3 at §3.2.3.
- 106 Still, it would have been nice to have been given a safe harbor, say if there were three or more other R&D poles that would create a presumption (at least) that the deal should be cleared. See, e.g., Arnold & Porter Comments, at ¶5, available at http://ec.europa.eu/competition/consultations/2009_horizontal_agreements/arnold_porter_en.pdf.
- 107 While the rationale for the Commission’s actions is not on the public record as such, the lead lawyer for DG Competition, Luc Gyselen (now with Arnold & Porter in Brussels), has written it up in *Competition in Innovation: A Novel Concept – The Case Law on Pharmaceuticals*, ON THE MERITS—CURRENT ISSUES IN COMPETITION LAW AND POLICY 41-42 (L.Hancher & P. Lugar, eds) (2005).
- 108 *Id.* In a private note to me, Gyselen describes the case as being the only one in his recollection where the Commission looked solely on the impact of the joint venture on innovation to conclude that Article 101(1) applied.
- 109 See 2004 *EU Horizontal Merger Guidelines* at ¶9, available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2004:031:0005:0018:EN:PDF>.
- 110 *Id.* at ¶38.
- 111 No divestiture of any human product was required in that case, so the discussion of the legal approaches was without prejudice to any result.
- 112 See, e.g., In re: Pfizer Inc., Case No. 001-0059 (2000), at <http://www.ftc.gov/os/caselist/c3957.shtm>, discussed earlier.
- 113 The well-known philosophical theory of Occam’s Razor counsels that where there are two approaches that lead to the same result, the simpler one is usually to be preferred. That certainly would seem to apply here. <http://www.xs4all.nl/~johanw/PhysFAQ/General/occam.html>.

- 114 See Carrier, *supra* note 4 at 417-419 (collecting sources); Gotts & Rapp, *supra* note 4 at 101.
- 115 The pressure to approve drugs to treat AIDS is a case in point, where a drug that looks possibly effective may well be fast tracked; see, e.g., Greenberg, *AIDS, Experimental Drug Approval, and the FDA New Drug Screening Process*, 3 LEGISLATION & PUBLIC POL'Y 295 (2000), available at http://law.nyu.edu/ecm_dlv2/groups/public/@nyu_law_website__journals__journal_of_legislation_and_public_policy/documents/documents/ecm_pro_060626.pdf; see also the general FDA Statement on Priority Review, available at <http://www.fda.gov/forconsumers/byaudience/forpatientadvocates/speedingaccesstoimportantnewtherapies/ucm128291.htm>.
- 116 Indeed that is the message of the 1995 IP Guidelines; IP Guidelines, *supra* note 3, §§ 2.0 and 2.1.
- 117 See Rosch Speech, *supra* note 1 at 13-14.
- 118 See <http://cat.inist.fr/?aModele=afficheN&cpsidt=1474187> for a brief overview.
- 119 See <http://www.thestreet.com/story/10781009/1/drug-firms-team-on-alzheimers.html>.
- 120 Rosch speech, *supra* note 1 at 21.
- 121 In that regard, the thesis here agrees with the observation of Timothy Muris in his article about bundled discounts: Many of the models and assumptions being applied in innovation market analysis not only lack empirical testing, but indeed are contradicted by the conduct and motivations in the real world. See Muris & Smith, *Antitrust and Bundled Discounts: An Experimental Analysis*, 75 ANTITRUST L.J. 399 (2008).



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Practical Aspects of Aftermarkets in European Competition Law

John Temple Lang
Cleary Gottlieb & Trinity
College

Practical Aspects of Aftermarkets in European Competition Law

*Dr. John Temple Lang**

An “aftermarket” is a market for the supply of products or services needed for or in connection with the use of a relatively long-lasting piece of equipment that has already been acquired. Aftermarkets give rise to several kinds of questions under competition law. Does a relevant market for competitive analysis consist of separate markets for primary and secondary products, or is it a market for “systems” consisting of both primary and secondary products? When, if at all, is the supplier of the primary product dominant in the aftermarket for products or services needed for use with its equipment? If it is dominant, what conduct may be an abuse prohibited by Article 102 (ex-82) of what is now the Treaty on the Functioning of the European Union? This article applies well-known general principles of competition law, along with case law, to answer these questions.

*Cleary Gottlieb Steen and Hamilton, Brussels and London; Professor, Trinity College, Dublin; Senior visiting Research Fellow, Oxford

I. Introduction: What the Key Issues Are

An “aftermarket” is a market for the supply of products or services needed for or in connection with the use of a relatively long-lasting piece of equipment that has already been acquired. Aftermarkets include, for example, repair and maintenance services, spare parts and consumables such as cartridges for printers, games for games consoles, coffee capsules for coffee machines, and replacement blades, e.g. for simple shaving razors or for complex drilling and earth moving equipment. They include computer software and hardware, and any situation in which consumers buy one piece of equipment (or one copy of a computer program) and afterwards buy more of the same equipment or program.¹

The word is also used, less precisely, to describe the supply of improvements and additional products and services that become available during the life of the equipment. This category can include software upgrades that may become available in the market, and may be convenient or necessary for smooth functionality of the original software licensed by the user. (Eventually a chain of aftermarkets may develop with the successive acquisition of the hardware device, plus the subsequent software, plus additional software updates and upgrades.)

It is convenient to refer to all these aftermarket products and services as “secondary products,” and to the products that were first acquired as “primary products.”²

Aftermarkets give rise to several kinds of questions under competition law. Does a relevant market for competitive analysis consist of separate markets for primary and secondary products, or is it a market for “systems” consisting of both primary and secondary products? When, if at all, is the supplier of the primary product dominant in the aftermarket for products or services needed for use with its equipment? If it is dominant, what conduct may be an abuse prohibited by Article 102 (ex-82) of what is now the Treaty on the Functioning of the European Union?

If the secondary products or services in the aftermarket can be used with all the primary capital equipment produced by different manufacturers, these questions do not usually arise. There are then two multi-brand markets, for the capital equipment and for the aftermarket products or services.³ This situation may arise as a result of standardization, or informally.

Competition problems may arise if the secondary products or services for use with each manufacturer’s primary product cannot be used with primary products produced by any other manufacturer. In this situation, the buyer of the primary

DOES A RELEVANT MARKET FOR COMPETITIVE ANALYSIS CONSIST OF SEPARATE MARKETS FOR PRIMARY AND SECONDARY PRODUCTS, OR IS IT A MARKET FOR “SYSTEMS” CONSISTING OF BOTH PRIMARY AND SECONDARY PRODUCTS?

product may need to buy the secondary products or services that it needs from the manufacturer of the primary product. It may then be eventually alleged that the manufacturer's prices are excessive, that the manufacturer is illegally tying the supply of the secondary products or services to the primary products, or the manufacturer is refusing to enable competing producers of the secondary products to supply them.⁴

In theory, according to the U.K. Office of Fair Trading, there may be either:⁵

- A “system market,” consisting of a single market for combinations of primary products and the secondary products (e.g. a market for all razors and replacement heads). This may be because buyers engage in “whole life costing,” taking into account the cost of the secondary products when they choose the primary product.
- “Multiple markets,” consisting of a market for the primary products, and separate markets for the secondary products compatible with each manufacturer's primary product (e.g. one market for all razors, individual markets for each type of replacement head).
- A “dual market,” consisting of a market for the primary products and a separate market of all the secondary products, if they are compatible with all the primary products.⁶

USING THIS TERMINOLOGY,
COMPETITION LAW PROBLEMS MAY
ARISE ONLY IN “MULTIPLE
MARKETS” IN WHICH EACH
OF THE SECONDARY PRODUCTS
IS COMPATIBLE WITH ONLY
ONE MANUFACTURER'S
PRIMARY PRODUCT.

Using this terminology, although the Office of Fair Trading does not say so, competition law problems may arise only in “multiple markets” in which each of the secondary products is compatible with only one manufacturer's primary product. It is only in such situations that there may be separate markets for the secondary products, and there may be dominant positions in those markets. However, even in what are described as “multiple markets,” the “market” for each manufacturer's secondary product is

linked with and depends on the demand for the primary product. In practice, therefore, such a “market” is not separate and is not a market in which the manufacturer could be dominant unless it is also dominant for the primary products.

There are no reported cases in Europe in which a manufacturer has been found dominant for its secondary products when it was not dominant for the primary products.⁷ In other words, whether one asks if there are separate markets or if there is dominance, one reaches the same conclusion when the competition for the primary product constrains the manufacturer of the secondary product. The criteria for market definition and for dominance in practice lead to the same result, whatever it is, in all cases.

II. Market Definition (Is the market for “systems” consisting of both the equipment and the consumables?) and Dominance (Can an undertaking not dominant in the primary market be dominant in the secondary market?)

A. THE GENERAL PRINCIPLES

In economic and legal theory, the relevant principles about market definition and dominance assessment in aftermarkets are now clear and well-established.

It is well-known that at least some potential buyers of capital equipment are likely to be aware of, or influenced in their choice by, the cost of the products and services in the aftermarket. If this is so, the test known in Europe as the *Pelikan-Kyocera* test⁸ is applied, and there is a single market for the “systems,” the combinations of the equipment and the products or services that will be needed when the equipment is being used.

In the *Pelikan-Kyocera* case, the Commission used the phrase “primary and secondary markets” but went on to conclude that the cost of the secondary products was taken into account by buyers of the primary products (and, therefore, although the Commission did not say so, that they were *not* separate markets). The Commission concluded that lack of dominance in the market for the primary products prevented dominance for the secondary products arising where both products are interrelated, as evidenced by:

1. Whether the consumer can make an informed choice including lifecycle pricing;
2. Whether the consumer is likely to make an informed choice;
3. Whether, if exploitation of the aftermarket occurred, a sufficient number of customers in the primary market would adapt their purchasing behavior at the level of the primary market, (that is, would switch and buy another manufacturer’s primary product); and
4. Whether such adaptation would take place within a reasonable period of time.

A more correct reference to a “single market” combining primary and related secondary products appears in the Commission’s recent guidelines on vertical restraints:⁹

“(91) Where a supplier produces both original equipment and the repair or replacement parts for this equipment, the supplier will often be the only or

the major supplier on the aftermarket for the repair and replacement parts. This may also arise where the supplier (OEM supplier) subcontracts the manufacturing of the repair or replacement parts. The relevant market for application of the Block Exemption Regulation may be the original equipment market including the spare parts or a separate original equipment market and after-market depending on the circumstances of the case, such as the effects of the restrictions involved, the lifetime of the equipment and importance of the repair or replacement costs. **In practice, the issue to decide is whether a significant proportion of buyers make their choice taking into account the lifetime costs of the product. If so, this indicates there is one market for the original equipment and spare parts combined.**" (emphasis supplied)

How far potential buyers are influenced will depend on many circumstances:¹⁰ the sophistication of at least some of the buyers; how expensive the equipment is; the relative costs of the equipment and the aftermarket products over the average life of the primary equipment; how long the equipment will be used; how much the buyers know or can find out about the prices in the aftermarket and take them into account upon time of purchasing the primary product (prediction of lifetime costs); whether the secondary products must be bought on a continuing basis (as in the case of consumables or routine maintenance) or only irregularly when repairs or improvements are needed (repetitive buyers compared with sporadic buyers in the aftermarket); whether there is a high proportion of potential (new) customers compared with current (old) customers; whether it is possible to discriminate in prices for secondary products between new and old customers; whether the equipment can be hired or must be purchased; what the total net cost of replacing the equipment with that of another manufacturer would be (switching costs, which may depend, among other things, on the second hand value of the equipment, and the cost of retraining, installation, and changing software); and how long it would take to switch. If the aftermarket includes consumables, the buyer may need to estimate how much of the consumables the buyer will use, and be able to make meaningful comparisons with other primary products.¹¹ Clearly, it is important, in the case of spare parts, that many of them are available from sources other than the manufacturer of the primary equipment.

It will be seen that almost all of these factors are questions of degree, and that even for a given manufacturer's equipment there may be a wide range of potential buyers, whose needs and interests may differ widely. If the equipment is relatively expensive, some buyers will certainly calculate the cost of consumables as well.

In these circumstances the manufacturer of the equipment might, if it can, seek to charge different prices to different types of buyers. If it cannot price-differentiate for the initial capital cost (as would usually be the position), the buy-

ers for whom the marginal cost of the systems are most important will keep the price of the whole system down, particularly if they have bargaining power, thereby benefiting the infra-marginal buyers. The key question is whether there are enough price-sensitive buyers to make a price increase unprofitable.¹² In fact it is common for equipment manufacturers to keep the price of consumables high relative to the price of the equipment, precisely because that enables the manufacturers to obtain greater revenues, overall, from buyers making the most use of the equipment; in effect, they succeed in price differentiating through the sales of the consumables.

There are so many factors that may need to be taken into consideration that it might sometimes be difficult to say with confidence (if it were necessary), what proportion of the buyers are significantly influenced by the cost of the secondary products, and therefore how the single market for the systems really works. It would not be necessary to show actual changes by buyers to another primary product, but discipline between primary and secondary markets can be concluded from showing the possibility of such changes.

It is useful to distinguish between: (i) secondary products (e.g. consumables and maintenance) which are likely to be needed in direct or known proportion to the extent of use; (ii) repair services and spare parts which may be required irregularly and unforeseeably; and (iii) improvements or extra features such as new games or altered software. Improvements may be optional at least for some customers, but the other secondary products are not. Repairs and spare parts may be needed only at long intervals, or as a result of accidents or mishandling. Some spare parts, however, wear out at a foreseeable rate and need to be replaced regularly, so are more like consumables. Because consumables are needed in proportion to the extent of use, the potential buyers will be more likely to be influenced by the lifetime costs when purchasing the primary product. However, the differences between the degrees of foreseeability of necessary items do not basically alter the competition law analysis. Improvements, which may be either necessary or optional, and which imply a dynamic market, raise rather different issues, and need separate consideration, as described below.

It is important to note that the buyers need not be able to calculate accurately how much they will spend in the aftermarket over the life of the equipment. It is enough if a significant proportion of them, especially those with bargaining power, estimate it as best they can, and take it into consideration when making their choice of equipment. It would be quite wrong to assume that an equipment

THERE ARE SO MANY FACTORS THAT MAY NEED TO BE TAKEN INTO CONSIDERATION THAT IT MIGHT SOMETIMES BE DIFFICULT TO SAY WITH CONFIDENCE, WHAT PROPORTION OF THE BUYERS ARE SIGNIFICANTLY INFLUENCED BY THE COST OF THE SECONDARY PRODUCTS, AND THEREFORE HOW THE SINGLE MARKET FOR THE SYSTEMS REALLY WORKS.

manufacturer is dominant for *e.g.*, repairs, merely because buyers do not know precisely how many repairs will be needed. If enough buyers are well enough informed about the probable cost of the whole “system” (primary and related secondary products), a significant increase in the price of secondary products or services would prove unprofitable overall, even if some buyers were unsophisticated or believed that they would use so few secondary products that the total cost would be unimportant. This is, in effect, the application of critical loss analysis by the manufacturer: The profits lost when well-informed buyers switch might be greater than the extra profit from higher charges paid by users who stay. The Commission seems not to have given enough attention to this point in the past.

It does not seem necessary to show how many customers would switch, or how soon, if the price of secondary products rose unduly. If prices have never risen unduly, it would be impossible to show the result empirically. It must be enough that it would, at some point, be possible and rational for at least some customers (presumably those buying the largest quantities of secondary products), to switch.

Many manufacturers of equipment practice “systems pricing,” that is, they charge relatively low prices for the equipment, in the hope or expectation of making a reasonable profit overall during the life of the equipment by charging appropriate prices for the secondary products. The fact that manufacturers do this is not evidence they believe that buyers are unaware of or uninfluenced by prices in the aftermarket. The manufacturers may know that buyers making little use of the equipment are likely to be attracted by a low initial cost, and are less concerned by the prices of the secondary products because they will not buy

INDEED, THE USE OF SYSTEM PRICING IS SO WIDESPREAD THAT IT IS HARD TO IMAGINE A BUSINESS BUYER SO NAIVE THAT HE WOULD BE UNAWARE OF THE POSSIBILITY THAT CONSUMABLES OR REPAIRS (SERVICES AND PARTS) WOULD BE PART OF THE COST OF USING THE EQUIPMENT.

very many. The manufacturers also must assume that buyers who will make much use of the equipment will be sophisticated enough to calculate carefully the overall cost during the life of the equipment and, may, anyway be attracted by the low initial cost for cash flow or tax reasons.¹³ Consumables are current expenses, but the cost of capital equipment may need to be depreciated over more than one tax year.

Indeed, the use of system pricing is so widespread that it is hard to imagine a business buyer so naive that he would be unaware of the possibility that consumables or repairs (services and parts) would be part of the cost of using the equipment, and its use might become expensive. Everyone who buys a photocopier or a printer, or indeed a car, knows that.

Further, in any given market, if one manufacturer uses “system pricing” all manufacturers normally are obliged to do so, because the attraction of low initial capital expenditure is considerable.

“Systems pricing” is not the only evidence that points to the conclusion that the relevant market is for “systems.” In addition to cost-of-ownership information and low switching costs, company reputation¹⁴ and other factors protecting existing customers against “installed-base opportunism” (that is, overcharging of existing customers for secondary products on the assumption that they are locked in and can be forced to pay), are also important. There are times when this overcharging might be tempting: the product is in a declining market; the firm is having trouble competing in the market; the products are marginally profitable or unprofitable on a life-cycle basis, the firm has a few other products whose goodwill would not be affected; and the firm is in financial distress or has a very high cost of capital.¹⁵ In all other cases, installed-base opportunism is unlikely. A company selling multiple product lines for the same customers will not engage in installed base opportunism for one product line, as this would harm sales of other product lines. A high degree of technical change leading to short life of equipment would also discourage exploitation of current customers who have to make new purchasing decisions as soon as their equipment becomes obsolete.

All these factors have to be considered when assessing whether a dominant position could exist for secondary products, or whether competition in the primary products is disciplining competition and behavior in the sale of secondary products. Installed base opportunism, if it occurred, would involve charging high prices to users, rather than refusing to sell secondary products to downstream competitors. A manufacturer could sell secondary products to competitors and charge them high prices.¹⁶ A refusal to sell secondary products to competitors is not evidence of installed-base opportunism.

Innovation as a key factor in competition in the primary products—as it typically happens in technological devices—may be also another factor to consider. If an increase in prices occurs for the secondary product, customers may have an additional incentive to switch to alternative primary products that offer the latest technology.

The conclusion of the above is clear, and complications are unnecessary. If there is competition for the primary products, it will, in practice, constrain the competition for the secondary products, even if the secondary products cost so little that buyers are not likely to estimate the total cost over the life of the primary product. Antitrust problems could arise only in the unusual cases in which a buyer cannot estimate the overall cost of the manufacturer’s system; costs for secondary products are not material; and switching costs for primary products are high.

IF THERE IS COMPETITION FOR THE PRIMARY PRODUCTS, IT WILL, IN PRACTICE, CONSTRAIN THE COMPETITION FOR THE SECONDARY PRODUCTS, EVEN IF THE SECONDARY PRODUCTS COST SO LITTLE THAT BUYERS ARE NOT LIKELY TO ESTIMATE THE TOTAL COST OVER THE LIFE OF THE PRIMARY PRODUCT.

B. THE CASE LAW OF THE EUROPEAN UNION COURTS ON AFTERMARKETS

The economic principles described above have been seldom considered by the European Courts. There are probably several reasons. First, the economic principles are likely to show, in practice, an absence of dominance in secondary products unless there is dominance for primary products, so the Courts have no reason to dispute the results when these principles have been applied by competition authorities. Second, like many principles that have never been litigated, they are so widely accepted that they could hardly be contested. A universally accepted principle may be more firmly established than one based on a single, perhaps controversial, judgment. Third, these economic principles are consistent with common sense and general experience—a useful basis for a rule of law. Nevertheless, the main cases concerning aftermarkets that have been decided by the Courts must be mentioned.

In *Hugin Kassaregister*¹⁷ the Court annulled the Commission's decision on the ground that there was no evidence that the refusal to supply spare parts needed for servicing mechanical cash registers had an effect on trade between Member States. The Court accepted, as a fact, that there was a specific demand for Hugin spare parts (parts for other brands did not fit Hugin machines), and that a manufacturer could be in a dominant position for the supply of its own spare parts. However, the Court did not consider the argument that Hugin was not in a dominant position in the equipment market for cash registers, and therefore could not be in a dominant position for spare parts. In other words, the *Hugin* judgment

IT SEEMS MOST UNLIKELY THAT THE COMMISSION WOULD REACH THE SAME CONCLUSIONS TODAY IF A CASE SIMILAR TO HUGIN WOULD ARISE.

does not mean that the Court rejected the economic principles explained here. It seems most unlikely that the Commission would reach the same conclusions today if a case similar to *Hugin* would arise.

In *Hilti*¹⁸ the Court found that Hilti was dominant for the equipment (cartridge-operated nail guns) and accepted that the company was also dominant for the consumables. This judgment is therefore entirely consistent with the economic principle outlined above.

*Tetra Pak*¹⁹ was a more complicated case, discussed below in connection with tying. *Tetra Pak* was found to be dominant on the market for aseptic packaging, but to have committed an abuse by seeking to monopolize the market for non-aseptic products, although the two kinds of products were not sufficiently interchangeable to form a single market.

The most recent judgment of the European General Court, concerning *Swiss Watches*, is discussed separately, below, in Section II (D).

C. EUROPEAN COMMISSION AND NATIONAL COMPETITION AUTHORITY CASES

It is useful at this point to summarize several of the key decisions of the Commission, as well as some other cases. It will be seen that a distinction has not always been clearly drawn between the question of whether there were separate markets for the primary and secondary products (rather than one market for “systems”) and the question of whether, if there were separate markets, the company in question could be dominant in the secondary market. The Commission has described primary and secondary products as if they were separate markets even when it went on to conclude that they were part of a single market. However, in spite of imprecise use of language, the principles are now clear.

All these cases involved complaints by competing producers of secondary products, not by customers.

As mentioned already, in *Pelikan/Kyocera*,²⁰ the Commission rejected the complaint of Pelikan, a German manufacturer of toner cartridges for printers, against Kyocera, a Japanese manufacturer of computer printers including toner cartridges for those printers. Pelikan’s complaint alleged a number of practices by Kyocera to drive Pelikan out of the toner market and accused Kyocera of abusing its dominant position in the secondary market. Kyocera was clearly not dominant in the primary market. There was no evidence of behavior that could be considered abusive. The Commission did not find that Kyocera had a dominant position in the market for consumables.

This finding was due to the features of the primary and secondary markets. Purchasers were well informed about the price charged for consumables and took this into account in their decision to buy a printer. “Total cost per page” was one of the criteria most commonly used by customers when choosing a printer, because life-cycle costs of consumables (mainly toner cartridges) represented a very high proportion of the value of a printer. Therefore, the Commission considered that if the prices of consumables of a particular brand were raised, consumers would have a strong incentive to buy another printer brand. In addition, there was no evidence that price discrimination between “old” or captive customers and new customers would be possible.

It is useful to compare *Pelikan/Kyocera* with the *Digital* case,²¹ in which the Commission obtained an undertaking from Digital on the assumption that Digital was dominant on the market for maintenance of its computers and software, but without making a formal decision regarding market definition, dominance assessment, or abusive behavior. The Commission said that switching hardware and software systems was slow, expensive, and difficult, and involved re-writing related systems. The prices of maintenance services were individually negotiated and confidential, and different terms could be given to different kinds of customers. Costs were not the primary consideration in choice of systems. Digital had a large base of customers that could not easily switch, and Digital

negotiated special terms with new customers.²² Because Digital gave an undertaking, the question whether the Commission was right to consider it dominant was not resolved. Evolution of servers from the proprietary systems in 1995 to today's open standard systems makes the *Digital* case less relevant for the purposes of market definition.

In *Infolab/Ricoh*²³ the Commission rejected a complaint of Info-Lab, a manufacturer of toner for photocopiers, against Ricoh, a photocopier manufacturer. Info-Lab alleged that Ricoh abused its dominant position on the market for toner cartridges compatible with certain Ricoh photocopiers and protected by Ricoh intellectual property rights by refusing to supply Info-Lab with empty toner cartridges, which would enable Info-Lab to compete with Ricoh in the sale of filled toner cartridges. Info-Lab claimed it was not possible to design a toner cartridge that would fit into the Ricoh machines without infringing Ricoh's intellectual property rights.

The Commission ruled that consumers were able to make informed choices when buying photocopiers and were, in practice, likely to do so. If Ricoh had tried to take advantage of its position by raising the price of cartridges, a sufficient number of consumers would switch their purchases of photocopiers, within a reasonable period. The Commission said that Ricoh was not dominant for photocopiers, and as the supply of cartridges was closely linked to photocopiers, it could not be dominant for cartridges.

In the *EFIM* case²⁴ the Commission explained its thinking in the two previous printers/photocopiers cases. These cases had raised the question whether a company could be dominant in the consumables market where it was not dominant in the upstream market for printers. If there was a close link between them, competition on the primary market could constrain companies' behavior on the secondary market. The Commission had found that Kyocera could not be dominant on the market for consumables because it was subject to intense competition on the primary market. For the same reasons, Ricoh could not be dominant for consumables. In the *EFIM* case, none of the companies was dominant for inkjet printers and, applying the same principles, the Commission concluded that none of them could be dominant for consumables. Even if the markets for primary and secondary products might be considered separate for some purposes, they were so closely linked that the relevant market should be regarded as a "printing systems market" comprising both the primary (printers) and secondary (consumables) products.

*ICL/Synstar*²⁵ was a U.K. Office of Fair Trading decision finding that ICL was not in a dominant position for the supply and maintenance of computer equipment with mainframe functionality. No separate market, or relevant secondary market, existed for hardware maintenance services for ICL mainframes, because of whole-life costing by purchasers of hardware. Buyers were sophisticated, the whole life cost of the secondary product was high relative to the price of the pri-

mary product, and price information was transparent since most purchases were by competitive tender. Product life cycles were getting shorter, and so the occasions for switching arose more often. Because there was no dominant position, ICL's refusal to supply certain diagnostic software for maintaining non-ICL hardware was lawful. In the single market for the primary and secondary products, ICL was not dominant.

A slightly different issue arose in *Novo Nordisk*.²⁶ Novo Nordisk, a leading insulin producer, introduced a method of insulin self-injection, the “insulin pen” system. Other companies’ pen delivery systems include various components (i.e. injection devices, cartridges containing the insulin dosage, and disposable needles) which were compatible with the Novo Nordisk system. The Commission found that Novo Nordisk abusively disclaimed liability for the malfunction of its pen products, or refused to guarantee such products, when they were used with the compatible components of other manufacturers (even where malfunctioning might not be due to the use of those components).²⁷ Following discussions with the Commission, Novo Nordisk agreed not to use disclaimers in such circumstances and, further, not to treat other manufacturers’ components as incompatible with its pen systems merely because of Nordisk’s inability to carry out the ancillary “function check” for its pen products. In this case it appears that Novo Nordisk was considered dominant for the primary product. So this case is consistent with the conclusion that competition law issues exist in practice in aftermarkets only in case of dominance for the primary products.

It will be seen that the revised *Pelikan/Kyocera* test (essentially, do enough primary products buyers take into account the cost of the secondary products?) is unquestioned and valid in both law and economics. The soundness of the test should not be questioned merely because some terminology has been used loosely. If at least a significant proportion of primary products buyers take into account the cost of the secondary products, then the two kinds of products are so related that they are not in separate “markets,” and the secondary products are certainly not in a “market” in which a manufacturer can be dominant unless it is dominant in the market for the primary products. Apart from luxury products, it is hard to imagine a situation in which buyers would not take the cost of consumables into account, so there probably are no cases in which the secondary products form a separate market.

IT WILL BE SEEN THAT THE REVISED *PELIKAN/KYOCERA* TEST (ESSENTIALLY, DO ENOUGH PRIMARY PRODUCTS BUYERS TAKE INTO ACCOUNT THE COST OF THE SECONDARY PRODUCTS?) IS UNQUESTIONED AND VALID IN BOTH LAW AND ECONOMICS.

D. AN EXCEPTIONAL CASE OF LUXURY PRODUCTS—SWISS WATCHMAKERS

A judgment of the European General Court²⁸ shows that separate dominant positions in aftermarkets are unusual, but not impossible. The case arose from a com-

plaint by a confederation of independent watch repairers. The Commission considered that there was no separate market in repair and maintenance services, but that the supply of those services was a feature of the luxury watch market, which is highly competitive. The confederation appealed against the rejection of the complaint.

The Court noted that the Commission said that the spare parts market for primary products of a particular brand may not be a separate market if either (1) the consumer can switch to spare parts of another manufacturer or (2) the consumer can switch to another primary product to avoid an increase in the price of spare parts. Importantly, the Court added that it would need to be shown that “in the event of a moderate and permanent increase in the price of secondary products, a sufficient number of consumers would switch to other primary or secondary products” so that such a price increase would be unprofitable.

A JUDGMENT OF THE EUROPEAN
GENERAL COURT SHOWS
THAT SEPARATE DOMINANT
POSITIONS IN AFTERMARKETS ARE
UNUSUAL, BUT NOT IMPOSSIBLE.

The Commission had initially considered that spare parts for luxury watches of different manufacturers were not substitutable for one another, and that the evidence subsequently showed they were not. Therefore, the key question was whether consumers would avoid increases in the prices of spare parts by switching to another primary product. The Court found that the total cost of repair and maintenance of luxury watches over a ten-year period is, for most models, less than 5 percent of the price of a new watch. Also, the price of spare parts is normally included in the cost of repair and maintenance, so that a moderate increase in the price of spare parts would be negligible in comparison with the price of a new luxury watch.

A purely theoretical possibility of switching is not enough. The definition of the relevant market is based on the concept that effective competition exists, which presupposes that a sufficient number of consumers would actually switch and that a moderate price increase would therefore be unprofitable. The mere possibility that the consumer can choose between several brands in the primary market is not enough to make the primary market and the after market a single market, unless that choice is made, in part, on the basis of the competitive conditions in the secondary market. Indeed, the Commission had stressed that the cost of repair and maintenance, including spare parts, was insignificant in comparison with the initial cost of a luxury watch, and that consumers did not consider the cost of after-sales servicing when choosing a watch. So the Commission had made a “manifest error” in believing that a moderate price increase would be unprofitable.

The Court added that the existence of a large manufacturer of components and spare parts for watches, including luxury watches, which did not itself produce complete watches and which was therefore not in the primary market, was a strong indication of the existence of a separate spare parts market.

Therefore, the Court said that, due to the Commission's manifest error, "it cannot be ruled out that it might have established" that manufacturers of luxury watches held a dominant position at least in certain ranges of their spare parts that constitute separate relevant markets.

Presumably there are other luxury markets, e.g. for luxury cars, in which the cost of spare parts would be so small in comparison with the initial cost of the primary product that consumers would be unlikely to switch in response to a moderate increase in the price of spare parts. It seems unlikely that consumers who buy luxury products consider the cost of after-sales servicing or spare parts, even if the products require them, as cars and watches do.

There is, therefore, an important distinction between spare parts that are needed only if there has been wear or breakage, which may be a very small percentage of the total costs of the product, and consumables that are needed whenever the primary product is used, which may amount to 50 percent or 70 percent of the total cost of the product over its lifetime of use, depending on how much it is used. When making their initial choices, buyers of primary products that require consumables (at least if the buyers are professionals) can be expected to take the cost of consumables into account, and so consumables markets are most unlikely to be separate markets. A sufficient proportion of buyers of products requiring consumables would ultimately switch in response to a moderate price increase, and make the increase unprofitable overall.

THERE IS, THEREFORE, AN IMPORTANT DISTINCTION BETWEEN SPARE PARTS THAT ARE NEEDED ONLY IF THERE HAS BEEN WEAR OR BREAKAGE, AND CONSUMABLES THAT ARE NEEDED WHENEVER THE PRIMARY PRODUCT IS USED.

Even if a market for spare parts, or for repair and maintenance services, is separate from the market for the primary products, it does not follow automatically that the manufacturers of the primary products have dominant positions. There might be competing producers of spare parts, as there were in the *Swiss Watchmakers'* case. The normal market analysis would then be necessary to measure the market power of the manufacturers of the primary products.

This judgment is consistent with the economic principles explained in this article, because it was a situation of low information of aftermarket costs at the time of purchasing, low costs for secondary products (in proportion to the cost of primary products), and high switching costs).

E. THE COMMISSION'S 2005 DISCUSSION PAPER ON ABUSE OF DOMINANT POSITIONS

The Commission's Discussion Paper on Article 82 (now 102) in 2005²⁹ discussed aftermarkets, although the Guidance Paper in 2009³⁰ did not. The Discussion

Paper suggested a classic two-step test in order to first define the relevant market and then to assess dominance.

The Commission stressed that:

“The strong position of the supplier on such [secondary] product markets may, however, not be indicative of the actual degree of market power of the supplier, since it may be constrained by competition in the primary market (...). In such a situation the supplier of the primary product cannot be said to be dominant on the aftermarket.”³¹

On market definition, the Discussion Paper said that the aftermarket consisting of the secondary products of one brand of primary product may not be the relevant product market if it is possible to switch to the secondary products of other manufacturers, or if the cost of switching to the primary product of another manufacturer would not be excessive. This is, of course, essentially the economic principle stated above. The Discussion Paper however went on to consider the situation in which there is a separate aftermarket consisting of the secondary products of a single manufacturer. Suppliers of other secondary products might have difficulty competing, either because of the confidential know-how or patent rights of the manufacturer in question, or they might fear retaliatory competition for their secondary products.

On dominance assessment, the Discussion Paper analyzed those situations where an aftermarket consisting of the secondary products of one brand of primary products has been (provisionally) found to constitute a relevant product “market,” saying, “a dominant position on such a market can only be established after analysis of the competition on both the aftermarket and the primary market.”

The Paper further said, “competition on the primary market does not protect customers who have already bought the primary product from being harmed if the supplier changes policy and raises prices or lowers quality after the customer bought the primary product.” If the supplier does change policy, the supplier “is no longer restrained in the aftermarket by the link with the primary market. The supplier may therefore be found to have a dominant position.” This is a curious comment, and seems plainly incorrect because it assumes that the customer is locked in and has no possibility of switching. Also, it underestimates the importance of new customers and of current customers who will need to replace equipment. The Paper, however, accepted that installed base opportunism can be

restrained by long-term service contracts or non-discrimination clauses, as well as by switching.

The Paper also said that it is easier to estimate the total cost of the system if the aftermarket was for consumables, the cost of which depended simply on the amount of use, than if the aftermarket was for repairs and spare parts; “if only professional buyers make (accurate) life cycle calculations, the supplier may still have substantial market power in the aftermarket vis-à-vis private customers.” This comment is also curious because it assumes, without saying so, that the supplier can differentiate in the price of the secondary products between professional and private buyers, which is not usually correct. It also assumes that the number of “professional” buyers who would switch, if price differentiation was not possible, would not be sufficient to make a price increase for secondary products unprofitable overall.

In the Discussion Paper the Commission went on to say that if a dominant position has been found on this basis, “the Commission presumes that it is abusive for the dominant company to reserve the aftermarket for itself by excluding competitors from that market” by tying or refusal to deal.³² The refusal might be a refusal to license intellectual property rights, a refusal to supply information needed to provide secondary products, or a refusal to supply spare parts. This statement is also unfortunate because it seems to assume that any conduct (even pro-competitive conduct such as low prices, innovative functionality, or additional value) that indirectly keeps a competitor out of the aftermarket is, or is likely to be, illegal. But any company, even a dominant one, has a duty to create competition only when the refusal to contract would be illegal for some specific reason. A refusal to help a competitor to enter a market is not necessarily illegal.

A REFUSAL TO HELP A
COMPETITOR TO ENTER
A MARKET IS NOT
NECESSARILY ILLEGAL.

These assumptions caused the Discussion Paper section on aftermarkets to be regarded as unsatisfactory because it assumed that (i) companies could sometimes be dominant in an aftermarket even if the primary market is competitive and (ii) that if a company is dominant for the primary products, it is normally illegal to keep competitors out of the aftermarket. Neither of those assumptions seems correct.

The Paper was also rightly criticized for a number of other conclusions. It drew a distinction between prior purchasers and future purchasers, because it said that competition in the equipment market does not protect those who have already bought the equipment. This disregards the possibility of switching, and ignores the fact that every purchasing company is also a potential future buyer, sooner or later. No primary products last forever. It assumed that the manufacturer could price differentiate between new and existing customers, which, in most situations involving consumables, is unlikely. It failed to say that, in response to an increase in the price of consumables, one crucial question is whether enough

buyers would switch to make the increase unprofitable for the manufacturer. Finally, the 2005 Paper disregarded the importance of company reputation and other factors protecting existing customers against “installed-base opportunism.”

It is therefore understandable that the section did not reappear in the Guidance Paper in 2009. (It may also have been omitted because the Commission concluded, correctly, that, in practice, aftermarket for spare parts, repairs and consumables very rarely cause genuine competition problems, for essentially the reasons given here.)

O'Donoghue & Padilla³³ also correctly say that the Commission's comment about abuse was stricter than the tests applied in *Hilti*,³⁴ *Tetra Pak II*,³⁵ and *Microsoft*.³⁶ Dominance is not illegal; there must be evidence of abuse. Tying is normally pro-competitive, so the Commission's comment in the Discussion Paper about abuse, which implies something close to a *per se* rule against tying in the aftermarket in the case of a company dominant in the primary market, was also inconsistent with what the Discussion Paper itself said about tying.

In fact, “installed base opportunism” would be damaging, and perhaps ultimately suicidal, for any company that practiced it. A sensible company would be unlikely to deliberately sacrifice its future customers, and its own future interests, for the inherently short-term opportunity to exploit its existing customer base. An equipment manufacturer cannot differentiate between those existing buyers who are likely to buy primary products again in the future and those who are not.

This not only makes opportunism unlikely, but it also means that if it were practiced, the company would be in difficulties of one or more of the kinds mentioned above, would probably have decided to leave the market, and so probably would no longer be dominant. Therefore, installed-base opportunism seems to be a largely theoretical problem, since it could only be short-term and, if significant, would lead to increased switching to other primary products and would do lasting damage to any company that tried it.

The Discussion Paper added, briefly, that there might be justification or efficiency defenses;³⁷ for example, to guarantee the quality and good usage of the secondary products, or to enable the dominant company to save production, distribution, or transaction costs. The Paper did not say so, but an equipment manufacturer may obtain large economies of scale if it can supply all of the consumables or spare parts. Clearly the reputation of the capital equipment depends on the consumables or other products or services in the aftermarket being satisfactory.

If dominance has not been shown, it is, of course, unnecessary to consider such questions as tying, bundling, and refusal to supply information, spare parts, or patent licenses, or to consider the possible justifications or efficiency defenses for them.

In practice it is difficult to visualize a situation in which all the conditions of dominance in a single-brand secondary market would be fulfilled, except when the manufacturer is also dominant in the market for the primary product or, perhaps, in the case of luxury consumer products.³⁸ Customers are never locked in forever, even when they need to buy consumables for the primary product. Manufacturers in practice can rarely if ever differentiate in price between existing and new buyers or between professional and private buyers of consumables, because they could not legally, or in practice, prevent arbitrage by contractual restrictions on resale or otherwise.³⁹ If the cost of the secondary product was too high in comparison with the price of the equipment, the customer would switch sooner or later. As already pointed out, there are no reported cases in Europe of dominance for secondary products without dominance for the primary products.

F. THE QUESTION OF CONSUMER HARM

Finally, it is necessary to consider the harm caused to consumers as a result of high prices in the aftermarket. Where, as is often the case, a manufacturer is “system pricing” (that is, charging low prices for the primary products and making profits primarily or only from sales of consumables), it would obviously be wrong to measure the cost of consumables to consumers without taking the cost of the primary products into account. Lower prices in the primary product make the products available to more consumers. A “system price” is related to the actual use of the device made by each consumer. The issue, again, is whether there is information to make informed decisions.

Overcharging for spare parts and supplying spare parts to competitors in the maintenance market are separate issues. Shapiro pointed out⁴⁰ that consumers would not be likely to benefit much even if manufacturers make their spare parts available to competitors. Manufacturers would price the spare parts to take into account lost profits in maintenance. Imposing a duty to supply parts would reduce the value of the manufacturer’s intellectual property rights. Consumers would still need to compare the total costs of each manufacturer’s system with the total costs of other manufacturers’ systems. Spare parts, not labor costs, are the largest element in the cost of maintenance. Also, maintenance and servicing are commodity services, and there is limited scope for competition in providing them.

U.S. Law

Although this article is about European law, it is useful to look briefly at U.S. law, since similar issues have arisen there. *Eastman Kodak v. Image Technical Services Inc.*⁴¹ is regarded as the leading case. In that case the Supreme Court held that summary judgment in favor of the defendant in aftermarket cases is generally not appropriate, even when the company in question has no market power in the primary products, because competition in the primary product does not necessarily prevent market power and anticompetitive conduct in the secondary products. Switching costs for users might be high relative to the exploitative price increase,

and the number of locked-in customers might be high relative to new customers. Lack of information might mean inadequate competition in the aftermarket, and create circumstances facilitating installed-base opportunism. Then opportunistic policy changes by manufacturers taking advantage of lock-in effects might be illegal monopolization.

Kodak was found to have market power in the high-end photocopiers market (and not only in the spare parts market), so it was not a case of dominance only in the secondary market. In addition, there was a lack of after-market information (which was limited to Kodak reports on micrographic service and declarations by Kodak that Total-Cost-of-Ownership estimates included only a few initial years while the equipment generally lasted longer), and there was also “lock-in” due to high switching costs.

In practice, since the *Kodak* judgment the lower U.S. courts have almost always found that either switching costs were not so high or shortage of information was not so serious that there was market power over the secondary products.⁴² In addition, they have usually held that there could be a breach of Section 2 of the Sherman Act only if the manufacturer had altered its pricing or other policy after buyers of the primary products were locked in.⁴³ They have also found that there was no market power either because other manufacturers’ secondary products were substitutable (i.e. that there was a “dual market” according to U.K. Office of Fair Trading terminology), or because users contracted for secondary products when they bought the primary product, and therefore knew the cost in advance.⁴⁴ Also, in some cases, buyers were able to protect themselves by long term contracts against price increases. Lower courts have also concluded that there could be no power in the aftermarket because the relevant market is for “systems,” and because there is competition for primary products that limits the scope for monopolization of the secondary products.⁴⁵

WHETHER THE CASES ARE
ANALYZED BY REFERENCE TO
MARKET DEFINITION OR TO MARKET
POWER, THE RESULT IS THE SAME.

Whether the cases are analyzed by reference to market definition or to market power, the result is the same. This, in fact, had been pointed out by the Supreme Court in *Kodak* (at page 470 note 15): “Whether considered in the conceptual category of “market definition” or “market power,” the ultimate enquiry is the same—whether competition in the equipment market will significantly restrain power in the service and parts markets”

Some U.S. courts have considered that there can be monopolization in the secondary products, if the high switching costs and inadequate-information conditions are fulfilled, even if there has been no change in the manufacturer’s policy. That policy might be illegal if it is unchanged. On this view, any policy change does not create market power, it merely takes advantage of it, and is evidence of it. However, if there has been no policy change, there is no injury (except when the original contract was unfavorable to the buyer, which is not an antitrust violation).

The almost unanimous refusal of lower U.S. courts to make findings in favor of plaintiffs on the basis of *Kodak*, and the large volume of economic and legal criticism of the Supreme Court's judgment, led Hovenkamp in 2001 to say that the *Kodak* judgment should be overruled,⁴⁶ and Goldfine & Vorrasi⁴⁷ in 2004 to say that overruling was unnecessary, because it was not applied.

It should be clear that both the law and economics are essentially the same in Europe and the United States. In short, in spite of the differences between Article 102 TFEU and Section 2 of the Sherman Act, essentially the same result has been reached in both jurisdictions: No competition issues arise in practice unless the company in question is dominant in the primary products.

III. Article 102: Analysis of Abusive Behavior Typically Alleged in Aftermarkets

As it has been explained above, Article 102 in practice applies to an aftermarket only if the manufacturer is dominant in the supply of the primary products. There may be dominance over secondary products even though there is no dominance over the primary products only in rare cases. Unless there is dominance, none of the possible kinds of infringement of Article 102 discussed below can arise.

A. TYING AND AFTERMARKETS

The Commission's Guidance document on exclusionary conduct⁴⁸ says that the Commission will normally take action when an undertaking is dominant in the tying market (the market for the primary products, if an aftermarket is involved) and:

1. The tying and the tied products are distinct products, and
2. The tying practice is likely to lead to anticompetitive foreclosure.

This assumes that customers are being made to buy the tied (secondary) products with the tying products, and it later mentions that there may be objective efficiency justifications for tying.

Putting aside the question of efficiency justification for the moment, and on the assumption that there is dominance in the primary (tying) market, are there separate products for the purposes of a tying analysis? Physically there are, and primary products and secondary products (such as consumables and spare parts) are not always sold as a package. The fact that most buyers take the total cost of consumables into account does not necessarily mean that they could not be considered separate for tying purposes.

Tying is discussed in the Commission's 2009 Guidance Paper on exclusionary abuses, but not aftermarkets. The Guidance Paper says that the Commission will act against tying when "the tying practice is likely to lead to anticompetitive foreclosure."⁴⁹ The Commission is, of course, correct to say that there is a fundamental distinction between "anticompetitive foreclosure" and foreclosure that can occur legitimately because the dominant company is lawfully selling better products at lower prices, and thereby excluding rivals from the market. However, this raises the difficulty that the Commission has not clearly defined "anticompetitive foreclosure" or said clearly how to distinguish it from lawful and desirable competition, in the contexts of tying, aftermarkets, or generally.

THE FACT THAT MOST BUYERS
TAKE THE TOTAL COST OF
CONSUMABLES INTO ACCOUNT
DOES NOT NECESSARILY MEAN THAT
THEY COULD NOT BE CONSIDERED
SEPARATE FOR TYING PURPOSES.

anticompetitive foreclosure" or said clearly how to distinguish it from lawful and desirable competition, in the contexts of tying, aftermarkets, or generally.

The Guidance Paper merely says that anticompetitive foreclosure due to tying is more likely when the dominant company's conduct is lasting, or the company is dominant for more

than one product. The Commission's other comments (tying may lead to too few customers buying the tied product from rivals and the price may rise; tying may prevent buyers from altering the proportions of the two products that they use; tying may evade regulation of the price of the tying product) do not seem relevant to aftermarkets, or to complaints that the dominant company has a monopoly of an aftermarket.

The Guidance Paper says that tying by a dominant company is illegal if it harms consumers by anticompetitive foreclosure of one of the markets, and if the two products are distinct. This, correctly, implies that "foreclosure" is not always harmful to competitors or anticompetitive. However, the Commission's other comments do not help to indicate when tying might be harmful or anticompetitive. It may perhaps be that, although the Court in *Microsoft*⁵⁰ avoided weighing up or balancing the anticompetitive effects and the supposed justification, some such test may be needed, in at least some tying and bundling cases.

The tying section of the Commission's Guidance paper⁵¹ reads:

"The undertaking should be dominant in the tying market, though not necessarily in the tied market. In bundling cases, the undertaking needs to be dominant in one of the bundled markets. In the special case of tying in aftermarkets, the condition is that the undertaking is dominant in the tying market and/or the tied after-market."

The Commission also says “Two products are distinct if, in the absence of tying or bundling, a substantial number of customers would purchase or would have purchased the tying product without also buying the tied product from the same supplier.” Both comments are open to criticism.

One difficulty of applying tying principles to supposed exclusionary abuses in aftermarkets can be simply stated: If there have never been competitors in the aftermarket, it is difficult if not impossible to estimate what proportion, if any, of the buyers of the equipment would have bought secondary products from the competitors, if they had been able to do so. Everything would depend on the price, and perhaps also the quality and other competitiveness factors of the competitors’ products (innovation, design, ease-of-use, etc.). As products in an aftermarket (secondary products) are physically distinct from the capital equipment (primary product), there may be two products (a prerequisite for a tying abuse), even if they are not separate markets. The question of what buyers would have done if they had the choice is relevant to assessing the possible harm to consumers.⁵² However, since spare parts in aftermarkets are by definition all the same, price would be the only reason why consumers would buy them from different sources.

The U.K. Office of Fair Trading Assessment of Individual Agreements and Conduct, after explaining that tying is a form of leverage, correctly says:

“Where two activities are complementary (the sale of a product and the provision of maintenance services associated with the products, for example), there may be grounds for including them within the same relevant product market, in which case the issue of leverage of market power, as such, would not arise.”⁵³

The Guidance Paper, of course, accepts that tying often leads to efficiencies in production or distribution (including economies of scale) that benefit consumers, reduce transaction costs, or combine two products usefully into a single product. All of these efficiencies (particularly the first) are likely to be found in aftermarkets, but they do not seem to arise in aftermarkets in any special or unique way. In general, one would expect the manufacturer of equipment often to have unbeatable economies of scale in the production of spare parts and of consumables; although, competitors, of course, would have none of the research and development costs of producing the primary equipment and consumables in the first place.⁵⁴ In fact, if the secondary products were considered as a separate market, it might be one in which something approaching a natural monopoly would exist.

In *Tetra Pak II*,⁵⁵ the Court said that “even where tied sales of two products are in accordance with commercial usage, or there is a natural tie between the two products in question, such sales may still constitute abuse within the meaning of Article [102] unless they are objectively justified.” The Court did not accept that tying was a common practice with non-aseptic packaging, and said that even if it was, that would not justify tying of aseptic packaging.

IT SEEMS CLEAR THAT A VAGUE
“NATURAL LINK” OR “NORMAL
COMMERCIAL USAGE” IS NOT
ENOUGH JUSTIFICATION.

It seems clear that a vague “natural link” or “normal commercial usage” is not enough justification. Otherwise, the principle stated in *Tetra Pak II* may depend on whether there are two separate relevant markets and the company is dominant in both markets; if so, it adds nothing to the basic principle described here. Presumably in this context the possible abuse would be anticompetitive foreclosure, contrary to Article 102(b), and the justification would be the advantages for consumers of the tying practice. If so, this would necessitate estimating the scope for improvements being made independently by competitors producing only the tied products (bearing in mind that a company producing the tied products may thereby ultimately become able to enter the market for the primary products also).⁵⁶ Since, by definition, competitors in the aftermarket are selling the same products as the manufacturer, the scope for competition is limited. It would also be necessary to estimate the efficiency justifications resulting from the tying, which in most cases would be economies of scale, safeguarding of quality, encouraging innovation or expansion of output, or, in some cases, (e.g. blades for drilling and earth moving equipment) safety considerations.

Even if it were correct to regard primary and secondary products as being in separate markets, they would often be so closely related that it could rarely be unjustifiable or illegal to sell them together. This is particularly true as system pricing makes primary products available cheaply to more users, and so expands output overall. The efficiency justifications concern the primary market as well as the secondary products, and the whole situation must be taken into account.

If economies of scale are significant, they might be reduced as a result of one or more competitors entering the secondary product market; for example, as producers of the spare parts. It might therefore be necessary to try to assess whether the additional competition that would result from the entry of one or two competitors would offset the reduction in the economies of scale of the dominant company. So it might be necessary to distinguish between economies of scale that are achieved at a level of production that several companies can expect to reach, and which would not increase if their production increases further, and economies of scale that improve indefinitely with higher and higher production.

If economies of scale were easily reached, rival producers of secondary products such as consumables and spare parts could achieve them. If they can be obtained

only at large scales, it is hard to see a justification for obliging the manufacturer to share them with its competitors. In all cases, doing so would have the undesirable effect, recognized in all duty-to-supply cases, of discouraging development of alternative aftermarket sources of supply, and making the aftermarket less competitive than it would otherwise be. If the secondary products are subject to intellectual property rights, such as patents or copyright, imposing a duty to supply would involve a compulsory intellectual property license.

It therefore seems that, although for purposes of tying analysis primary and secondary products may be separate products, even in a market for systems, in practice dominance is unlikely, and tying is likely to be justified by efficiencies. Competition authorities wishing to encourage competition in secondary products need to be careful not to discourage innovation in the primary products.

COMPETITION AUTHORITIES
WISHING TO ENCOURAGE
COMPETITION IN SECONDARY
PRODUCTS NEED TO BE
CAREFUL NOT TO DISCOURAGE
INNOVATION IN THE
PRIMARY PRODUCTS.

B. REFUSAL TO SUPPLY OR LICENSE NECESSITIES FOR AFTERMARKETS TO COMPETITORS UNDER ARTICLE 102 (EX-82)⁵⁷

The Commission's Guidance Paper on exclusionary abuses (in effect, foreclosure) includes a section on refusals to supply competitors.⁵⁸ It merely says that a refusal case is an "enforcement priority" for the Commission if:

1. The refusal concerns a product objectively necessary to compete effectively on a downstream market;
2. The refusal is likely to lead to elimination of effective competition on the downstream market; and
3. The refusal is likely to lead to harm to consumers.

This may give the impression that a duty to contract will often arise in refusal to contract cases, which is most unlikely to be correct. The Commission cannot change the law merely by failing to mention principles stated or implicit in the case law of the EU Courts.

In aftermarket cases (e.g. refusal to supply spare-parts to third parties to enable them to compete on maintenance services), there does not seem to be any reason for a stricter rule on refusal to contract than in other cases. So it is appropriate to consider the limiting principles that apply generally. It will be seen that, in practice, in most aftermarket cases no duty to contract would be likely to arise, even in the unusual situation in which there is a dominant position for secondary products. In theory, a duty, if one arises, might be either to supply spare parts or software or to license the patents to allow their production.

Only if all the following conditions are fulfilled, a refusal to contract may be illegal under Article 102 because it excludes or forecloses a competitor from the aftermarket. Because of the risk in all duty-to-contract cases of discouraging innovation and pro-competitive investment, it is necessary to consider all the conditions and limiting principles, including those that are unfortunately not stated in the Commission's Guidance. They seem to be as follows:

- A duty to contract is exceptional, and is not a usual or normal obligation. There is no duty to contract merely because, without a contract, a competitor cannot enter the market. Such a duty would imply that a dominant company always has an obligation to create competition in a downstream market, even if it had committed no abuse, which would be absurd.⁵⁹ The mere failure to help a competitor to create competition is not, in itself, an abuse.
- There is a duty to contract only if refusal is, or is linked to, an illegal abuse for some specific and identifiable reason, and not merely because a contract would lead to more competition in the short term in the aftermarket. Merely adding one competitor, or making a competitor more competitive, is not a sufficient reason for imposing a duty to contract. Article 102 can be applied only if something illegal has been done. The fact that there would otherwise be a monopoly does not lead automatically to an abuse or automatically create a duty to contract with a competitor; that would be regulation, not competition law. The abuse might be discrimination (if a contract had been made previously in an "equivalent" situation); depriving consumers of a new kind of product for which there is a clear and unsatisfied demand (not merely a copy or another version of the same product); or some other significant action "limiting" the production, marketing, or technical development of a competitor to the prejudice of consumers, contrary to Article 102(b). Refusing to give a competitor a competitive advantage is not the same as "limiting" its possibilities. "Limiting" another's possibilities means creating a handicap or a difficulty to which the competitor would not otherwise be exposed. As a result, there is no duty to supply e.g. spare parts for maintenance unless a dominant company has done something identifiable which has created a difficulty for third-party maintenance firms (even if all the other conditions for a duty to supply are fulfilled).
- As in all cases under Article 102, there must be proof of harm to consumers, as well as exclusionary or anticompetitive foreclosure.⁶⁰
- As well as dominance upstream, there apparently must also be dominance, or at least market power, in the downstream aftermarket.⁶¹
- In any refusal to supply case, there are two markets: an upstream market for the product or service that is an essential input (e.g. spare parts), and the downstream aftermarket in which the complainant wants to use the input (e.g. for maintenance services). The upstream

market may be “potential,” that is, the dominant company may never have sold the product or service in question to anyone. But the fact that some other company wants the input in question is not enough to create a “potential” market. The mere existence of demand cannot create a duty to supply. That would mean that there would be a greater duty to share more valuable inputs, which would be anticompetitive. There is no general duty to provide an input in whatever form or at whatever stage it would be convenient for each individual complainant to obtain it. So there can only be a duty if the dominant company has previously sold,⁶² or if it would be economically rational for it to do so.

- There must be sufficient scope for added value competition in the downstream market. If there is no such scope, competition and consumers cannot be harmed by the refusal. There is no duty to supply if the competitor is going to provide only copies of the dominant company’s secondary products, or for simple distribution and resale of e.g., spare parts or identical consumables without any added-value services. There cannot be a duty to supply the dominant company’s final product to its competitors, which would be the effect if there was a duty to supply merely for distribution purposes.⁶³ This means that, in practice, there cannot be a duty to supply consumables if there is no scope for significant added value competition using them in maintenance services.
- The Court in *Microsoft* (a case in which there was dominance in the upstream market) said that the test under the *Magill* judgment (that there must be a new kind of product for which there is a clear and unsatisfied demand), is only one example of the possible harm to consumers which is required by Article 102(b).⁶⁴ Article 102(b) says that abuse may consist of “limiting production, markets or technical development to the prejudice of consumers,” and the EU Courts have held that this prohibits conduct limiting the production, markets, or technical development of competitors of the dominant enterprise.⁶⁵ But the mere fact that competitors’ production, marketing, or technical development is limited is not enough in itself to cause harm to consumers. If competitors will only copy the dominant company’s product or service, even with minor improvements, there is no justification for a duty to supply under Article 102(b). Copying is not added-value competition. Any slight short-term benefit to consumers would probably not outweigh the harm done to dominant companies’ incentives to invest.⁶⁶ It is not clear whether the scope for competition in maintenance in aftermarket would be sufficient for this purpose.
- The Courts said in *Bronner*⁶⁷ and *IMS Health* that a refusal to contract is illegal if it eliminates “all” competition in the downstream market. If the company seeking the contract is the only competitor, or if all other companies are being treated in the same way, the refusal to con-

tract eliminates all competition. If the dominant enterprise has already contracted with other companies, or if there are other companies that do not need to rely on the dominant enterprise, competition is not eliminated, and the refusal does not restrict competition. But if the only competitor is inefficient, there might still be a duty to contract if all the other conditions are fulfilled.

- In *Bronner*, the Court stressed⁶⁸ that there is no duty to supply if alternative solutions are available “even though they may be less advantageous,” and if there are no “technical, legal or even economic obstacles capable of making it impossible, or even unreasonably difficult,” for any competitor, alone or in cooperation with others, to establish its own solution. The fact that a small competitor would not find an alternative economically viable is not enough to create a duty to supply. It would be necessary “at the very least” to show that no alternative, even one involving several competitors, could achieve a comparable scale, and that economies of scale were important.
- Even when a dominant company has already made contracts, it is not necessarily discriminatory for it to refuse to enter into other similar contracts with “equivalent” parties. The refusal might have no effect on competition or do no harm to consumers (for example, if the claimant was plainly inefficient or if there were plenty of competitors already), or the refusal might be justified for some reason.
- If there is dominance, and an illegal refusal to contract is said to be justified by efficiencies, but the efficiencies could be obtained without causing the exclusionary effects, there may be a duty to avoid those effects. (For example, a genuine improvement in one of two products that must work together may make it incompatible with competitors’ versions of the other product. In this situation a dominant company may have a duty to provide the information needed by competitors for interoperability. The question of information about improvements is discussed below.)
- Since the mere refusal to license an intellectual property right or to supply a product cannot be, in itself, an exclusionary abuse, there must always be some other identifiable abuse, for which a duty to contract or to license is the appropriate remedy.⁶⁹ The fact that the intellectual property right enables the dominant company to monopolize the downstream market is not enough, in itself, to constitute an abuse.

In conclusion, the requirement of scope for added value competition, and the rule that mere copying is not enough to create a duty to contract, show that in most aftermarkets, because the proposed competitors merely want to copy what the equipment manufacturer is supplying, there is no duty to contract even if the manufacturer were shown to be dominant for the secondary product or service.

Determining the Lawful Price for Secondary Products: The Difficulty If a Remedy Were Needed

One further, important point needs to be made. Aftermarket controversies always involve either a refusal to supply competitors or setting a high price for the secondary products (or both). In either case, if an infringement were found, the competition authority or court would have to decide at what price the secondary products could lawfully be sold under Article 102. The European Commission has always avoided attempting to fix such a price.

Determining the “correct” price for products for which there is *ex hypothesi* no substitute is not really possible, in particular because it would be impossible to decide the “correct” (i.e. maximum lawful) price for the manufacturer’s secondary product without taking carefully into account its prices for its primary product, and deciding different “correct” prices for different categories of users. For example, users can be categorized based on intensity of use, or whether that use is private or commercial.

There are other factors that would have to be considered. If the secondary products were patented, the suggested remedy would also involve a compulsory license of the patents, and the “correct” royalty rate would need to be determined. A competition authority could not usefully or effectively order a dominant manufacturer to supply its competitors with spare parts without also determining the relationship between the price to competitors and the price to consumers. This would raise margin squeeze issues.⁷⁰ Competitors could benefit from a supply of spare parts only if their maintenance operations were at least as efficient as those of the manufacturer, which seems unlikely to occur often.

The difficulties of devising an efficient remedy are a further reason for extreme caution before using competition law in aftermarket cases even when it might seem to be justified; in particular, if the input in question has never been marketed so there is no empirical evidence of what an appropriate price might be.

DETERMINING THE “CORRECT”
PRICE FOR PRODUCTS FOR
WHICH THERE IS *EX HYPOTHESI*
NO SUBSTITUTE IS NOT
REALLY POSSIBLE.

C. EXCESSIVE PRICES FOR SPARE PARTS AND CONSUMABLES

Even if there is in some sense a market for the primary products, and a separate market for the secondary product associated with each primary product, it is suggested that a company that is not dominant for the primary product is rarely, if ever, dominant for secondary products consisting of spare parts and consumables for its own equipment (except in the unusual case of luxury products). Even if other manufacturers’ products do not work with its equipment, in practice competition is likely to be for the “system,” *i.e.*, the combination of equipment, spare parts, and consumables, especially where switching costs from one system to another are relatively low.

Where that view is applicable, the possibility that the prices for spare parts or consumables (as distinct from the prices for the system as a whole) would be so high as to infringe Art. 102(a) rarely arises in practice. If the company in question is dominant in the market for systems, it would be the overall price for the system that would need to be considered. If, as is usually the case with system pricing, the price of the equipment is relatively low, *i.e.*, not much above cost, that would have to be offset against apparently high prices for the secondary products (spare parts and consumables).

The question whether the prices of those secondary products (spare parts or consumables) might infringe Article 102(a) can therefore be dealt with briefly,⁷¹ keeping the following points in mind:

- System pricing based on a low profit margin for main equipment and a higher profit margin for consumables makes new primary products available to a larger number of new consumers than would otherwise have access to the equipment. It also makes it possible to have a wider choice of different “systems” and “pay-per-use” solutions for different types of customers. System pricing, in other words, is pro-competitive and expands overall output because it allows users a greater choice about the extent and perhaps the nature of their use.
- This situation does not cause consumer harm, as future revenues in the aftermarket are likely to be already discounted in the primary market. A range of prices for consumables also facilitates legitimate (and pro-competitive) price discrimination
- Spare parts for long-lasting equipment must normally be kept in stock, and if there are many parts, the cost of keeping a complete stock must be taken into account. Spare parts may need to be kept in stock for a considerable time after the equipment in question is no longer sold, if the parts for the latest equipment do not fit earlier models. (If the “spare part” is software, the problem of stocks does not usually arise.)⁷²
- It may become no longer economic to produce and stock parts or consumables for obsolete or obsolescent equipment, so the manufacturer may then license other companies if they wish to do so.
- The manufacturer must continue to bear the research and development costs for both future equipment and secondary products. These costs must be financed somehow. A manufacturer is free to spread the financing of its R&D costs across the whole range of its products as it wishes, and if much of its revenue comes from the aftermarket, it is entitled to attribute much of its R&D to that market, even if the R&D concerns the equipment rather the consumables (which is not necessarily the case).
- Apparently high prices for consumables may be an alternative to leasing out the primary products and charging royalties based on the

extent of use, which would be a legitimate alternative strategy for a manufacturer to adopt.

IV. Some Additional Issues Concerning Aftermarkets

A. IMPROVEMENTS IN AFTERMARKETS—THE MICROSOFT JUDGMENT⁷³

So far this paper has been concerned with aftermarkets in which the products or services were unavoidably necessary for users. However, there is at least one other situation, sometimes described as an aftermarket, to which rather different principles apply. This arises where the dominant manufacturer of the primary or capital equipment improves it in some respect, or offers additional equipment or functions or new video games or other complementary products. The question then is whether it is obliged to give its downstream competitors enough information about the improvement (or to provide the additional equipment) to adapt their aftermarket or downstream products or services to the improved product, or to enable them to produce downstream products or services that imitate it or are adapted to it. It has been suggested that providing information in this context is analogous to providing other things that are necessary for a competitor entering an aftermarket. However, the other situations discussed above arise primarily in static markets, but improvements by definition arise in dynamic markets, and for this reason at least may raise additional issues.

HOWEVER, THE OTHER SITUATIONS DISCUSSED ABOVE ARISE PRIMARILY IN STATIC MARKETS, BUT IMPROVEMENTS BY DEFINITION ARISE IN DYNAMIC MARKETS, AND FOR THIS REASON AT LEAST MAY RAISE ADDITIONAL ISSUES.

Another even more important distinction is that even existing users of the primary products usually do not need the improvements or additions, so for that reason they are not analogous to consumable secondary products or maintenance, which are unavoidable. Improvements are not consumables; being not necessary, they are optional. A manufacturer's market power, therefore, is much less in the case of optional improvements and additions. Buyers of the primary products make overcharging complaints about consumables, but complaints about refusal to provide improvements are typically made by competitors producing secondary products. These latter complaints therefore raise questions about the circumstances in which, under EU law, a dominant company may have a legal duty to supply or license its competitors, as discussed above.

The issue, as in the case of unavoidable necessities, usually is whether the refusal to give information about the improvement or to license the relevant technologies deprives consumers of a new kind of product for which there is a clear and unsatisfied demand. This is not likely to be true in the case of an improvement or addition, because by definition a new or at least an improved

product is being provided. So the key issue is likely to be whether the refusal to supply the improvement limits the possibilities of competitors contrary to Article 102(b), and so deprives consumers of a clearly proved probability of substantially improved products.

The Commission said in the Guidance Paper⁷⁴ that there is a duty to contract if the product is “objectively necessary to be able to compete effectively” and the refusal is likely to lead to “the elimination of effective competition.” But there is never a duty to make the competitor’s product better or more profitable, or to put it on equal terms, even when it is clearly not as good or as profitable as the dominant company’s product, unless the disadvantage of the competitor is due to some conduct of the dominant company,⁷⁵ as was found in the *Microsoft* case.

A more precisely and correctly expressed principle based on Article 102(b) would be as follows. A refusal to provide the information about an improvement or to provide add-on equipment that is needed by competitors downstream (including refusal to provide a license of intellectual property rights) may be an abuse, because it interferes with the dynamic process of competition, even if all the conditions listed above are fulfilled, only if:

1. Sufficient harm to consumers is shown;⁷⁶
2. The refusal will eliminate or permanently handicap competition and create or maintain dominance in a new or developing market for a new or substantially improved product that competitors are producing, (or would produce, if the evidence that they would do so is strong enough), and would continue to be under competitive pressure to produce.⁷⁷ The new or improved product would be a downstream or secondary product; and
3. The duty to contract would provide an essential input otherwise unobtainable without giving horizontal or direct competitors all, or most, of the dominant company’s competitive advantage or depriving it of the incentive to invest further. (As long as the dominant company exclusively retains its main competitive advantage, it has an incentive to invest. If its main competitive advantage has to be shared, the scope for added value competition would be reduced or eliminated and, even more seriously, the competition law rule would end the dominance, which EU law has no power to do.)

A principle about refusal to contract on these lines may be deduced from the *Microsoft* judgment, without listing all the special features of that case which, cumulatively, led to the conclusion that its conduct was illegal.⁷⁸ It seems likely that the *Microsoft* case will come to be seen as relatively unusual, and not as a forerunner to a large number of duty-to-contract cases in high technology industries. A principle on these lines would also largely avoid the need for balancing or weighing up benefits against disadvantages, which is difficult and unsatisfac-

tory, and which the Court in *Microsoft* did not do.⁷⁹ The limiting principles and conditions discussed above in connection with duties to supply are applicable.

B. THE IMPLICATIONS OF THE COMMISSION'S PHARMACEUTICAL ENQUIRY REPORT

In the preliminary report of its Pharmaceutical Enquiry,⁸⁰ the European Commission seemed to regard as probably illegal a wide variety of practices in connection with patents that are, at least in most cases, entirely lawful. This would be relevant to this paper if the Commission still believes, as it appeared to do at the time of the preliminary Pharmaceutical Report, that it is likely to be illegal for a company dominant for the primary product to use patent rights to exclude competition in the secondary products.

This uncertainty is another consequence of the fact that the Commission's Discussion and Guidance Papers do not comprehensively define exclusionary conduct. Therefore, they provide no sufficient test or criterion for deciding whether conduct is contrary to Article 102 TFEU if it does not fall into one of the well-recognized types of exclusionary abuse,⁸¹ patent practices do not.

So the result of the omission of an aftermarkets section from the Guidance Paper, and the aggressive statements made in connection with the preliminary pharmaceutical report, is to leave the Commission's view of the relevant principles in a state of uncertainty. The Commission may, in its Guidance Paper on exclusionary abuses and in its final report of the pharmaceutical enquiry,⁸² have arrived implicitly at a position less strict than it implied previously. Probably it has done so, but it would be helpful if this was confirmed.

C. NON-COMPETITION LAW ISSUES CONCERNING AFTERMARKETS

It may be useful to mention several other legal questions, not arising directly under European competition law, which can arise in connection with aftermarkets.

One question is whether, and how far, national patent laws in Europe allow a patent owner to sell a patented product and to limit how the purchaser or its customers can use the product. This was the issue in the U.S. Supreme Court in *Quanta Computer, Inc. v. LG Electronics, Inc.*⁸³ It is not clear how far the national laws of the EU Member States answer this question. This issue would arise if the equipment manufacturer sold patented consumables or spare parts but did not license the patents for use with competitors' refills or for re-use by competitors. European Union law does not answer the question either, because EU competition law does not prohibit a patent owner from giving a limited field of use license. (If the restriction was purely contractual, as it would be if no patent was involved, Article 101, ex-81, might apply.) Since the key question is whether the restriction would limit use by an indirect buyer, the question seems to be governed by national patent law, and not by European competition law.

ONE QUESTION IS WHETHER, AND HOW FAR, NATIONAL PATENT LAWS IN EUROPE ALLOW A PATENT OWNER TO SELL A PATENTED PRODUCT AND TO LIMIT HOW THE PURCHASER OR ITS CUSTOMERS CAN USE THE PRODUCT.

Other kinds of conduct have been said to raise issues under European competition law, some of which seem similar to questions arising from the Commission's Pharmaceutical Enquiry Reports. Apart from what is said to be manipulation of the technical features of the equipment or of the products in the aftermarket, the main issues concern intellectual property rights.

Clearly, in principle, even a dominant equipment manufacturer may obtain patents and exercise patent or other industrial property rights. Obtaining patents is pro-competitive, and even a dominant company is allowed to compete aggressively within whatever limits are imposed by competition law. So the mere fact that an equipment manufacturer has patented the design or production process for its consumables or spare parts cannot be contrary to Article 102. This is so even though it might make it more difficult for a competitor in the aftermarket to sell consumables to be used with the equipment of the manufacturer in question.

Clearly, also, a dominant company is free in principle to improve either the capital equipment that it sells or the secondary products, such as consumables or spare parts, to be used with it, or both. This is true even if one consequence of the improvement is to make the improved system incompatible with competitors' consumables or spare parts. However, if a company dominant in a "system market" changes the system only to make it incompatible with competitors' products, and provides no improvement (such as enhanced functionality or reduced production or distribution costs) but only has the effect of foreclosing competitors, that would likely be conduct limiting the production, markets, or technical development of competitors, and would be contrary to Article 102(b) if harm to consumers resulted.⁸⁴

V. Conclusions

The principle is well-established in law and economics that a supplier of aftermarket products is not dominant for them if a sufficient proportion of the buyers of the primary products are significantly influenced by the cost of the aftermarket products (such as consumables or maintenance services), because there is then a single market for systems which includes both primary and aftermarket products. Low switching costs, company reputation, and other factors protecting existing customers against "installed-base opportunism" are also important. If there may be cases in which a separate market is exceptionally defined for those secondary products, competition in the primary market is likely to discipline behavior in the secondary market. Except in the case of luxury products, it is difficult, if not impossible, to visualize a situation in which the buyers of the pri-

mary product would not be influenced by the cost of secondary products such as consumables, unless the prices of both (e.g., razors and razor blades) or of the secondary product (e.g. printer and A4 paper) are trivial, since the cost of consumables can always be ascertained and will be certainly ascertained by a sufficient proportion of end-users.

The same applies to other aftermarkets that involve expenditure in proportion to the use of the products (e.g. maintenance agreements).

So genuine cases of dominance, and therefore of possible abuse of dominance, are not likely to arise in those aftermarkets either in EU or in U.S. law unless the company in question is dominant in the primary market. This general conclusion is now well-established, and should greatly simplify assessment of aftermarket cases.

The picture might be less clear if genuine information problems exist in the aftermarket, e.g., if the aftermarket cost was unrelated to use but arose exceptionally or irregularly, due to accidents.

An alternative argument might be based on a supposed illegal refusal to supply competitors with inputs that are essential for entry into the aftermarket. However, there is no duty to grant a license or to supply goods or services merely to enable a competitor to offer essentially the same kind of products or services as the dominant company is already supplying. There would be a duty to supply only on the basis of a relatively narrow principle based on special circumstances similar to those of the *Microsoft* judgment. ▼

1 Carl Shapiro, *Aftermarkets and Consumer Welfare: Making sense of Kodak*, 63 ANTITRUST L.J. 483-511 (1995); he comments that "aftermarkets are ubiquitous."

2 Some documents from competition authorities provide additional guidance on what aftermarkets are:

DG Competition discussion paper on the application of Article 82 of the Treaty to exclusionary abuses (Dec. 2005): 243.

Aftermarkets are also sometimes called "secondary markets." Such markets comprise complementary products (or "secondary products") that are purchased after the purchase of another product (the "primary product") to which it relates. Standard examples include after sales services and spare parts for durable goods, as well as consumables such as ink cartridges and toner for printers and photocopiers. However, also upgrades of computer software may be considered aftermarkets.

The Commission's Notice on the definition of relevant market (1977) reads (¶56):

There are certain areas where the application of the principles stated above has to be undertaken with care. This is the case when considering primary and secondary markets, in particular, when the behaviour of undertakings at a point in time has to be analysed pursuant to Article [102]. The method of defining markets in these cases is the same, i.e. assessing the responses of customers based on their purchasing decisions to relative price changes, but taking into account as well, constraints on substitution imposed by conditions in the connected markets. A narrow definition of market for secondary products, for instance, spare parts, may result when compatibility with

the primary product is important. Problems of finding compatible secondary products together with the existence of high prices and a long lifetime of the primary products may render relative price increases of secondary products profitable. A different market definition may result if significant substitution between secondary products is possible or if the characteristics of the primary products make quick and direct consumer responses to relative price increases of the secondary products feasible.

UK Office of Fair Trading—“Market Definition” Guidelines (Dec. 2004):

An aftermarket is a market for a secondary product, that is, a product which is purchased only as a result of buying a primary product. For example, a customer would purchase a printer cartridge (a secondary product) only for use with a printer (the primary product). Another example is replacement heads for razors (the secondary product) and razors (the primary product).

- 3 See OFT Market Definition paper, *Id.* ¶¶ 6.2 and 6.4 (2004): “A dual market definition is appropriate where secondary products are compatible with all primary products (and perceived to be so by customers).”
- 4 See DG Competition Discussion Paper, *supra* note 2, ¶ 244:

Aftermarkets typically appear in competition cases when they are “proprietary”, that is, when they are brand-specific in that secondary products that can be used with one brand of primary product cannot be used with another brand of primary product, although the primary products themselves are substitutes. The contentious issue is often that the supplier of a primary product attempts to reserve the secondary market for itself.
- 5 See OFT Market Definition paper, *supra* note 2, ¶ 6.2.
- 6 Using these definitions, “Systems markets” and “multiple markets” are not mutually exclusive. The buyers of the primary products may take into account the probable cost of the secondary products over the life of the primary product, whether or not the secondary products are available only from one source. Even if it were difficult to estimate in advance the total cost of secondary products, it is not easy to imagine why the high cost of secondary products would not become known, at least to a significant proportion of buyers, and injure the reputation of the primary products with which they are compatible.
- 7 In the Swiss Watchmakers case (Case T-427/08, *CEAHR v. Commission*, Judgment dated December 15, 2010) the Court merely said that it could not be excluded that there was a dominant position. It is well recognized that products in clearly separate “markets” may constrain the ability of producers of each product to raise prices. Household fuels such as electricity, gas, coal, and wood are in separate markets for most purposes, but they clearly constrain the ability of producers to raise prices in the other markets. “The objective of defining a market in both its product and geographic dimensions is to identify those actual competitors of the undertakings involved that are capable of constraining those undertakings’ behaviour”: Commission Notice on definition of relevant market, OJ No. C-372/5, ¶2 (Dec.9,1997). At ¶56 of the Notice the Commission said:

There are certain areas where the application of the principles stated above has to be undertaken with care. This is the case when considering primary and secondary markets, in particular, when the behaviour of undertakings at a point in time has to be analysed pursuant to Article [102]. The method of defining markets in these cases is the same, i.e. assessing the responses of customers based on their purchasing decisions to relative price changes, but taking into account as well, constraints on substitution imposed by conditions in the connected markets. A narrow definition of market for secondary products, for instance, spare parts, may result when compatibility with the primary product is important. Problems of finding compatible secondary products, together with the existence of high prices and a long lifetime of the primary products

may render relative price increases of secondary products profitable. A different market definition may result if significant substitution between secondary products is possible or if the characteristics of the primary products make quick and direct consumer responses to relative price increases of the secondary products feasible.

- 8 Commission, XXVTH REPORT ON COMPETITION POLICY, pp. 41-42 and 140 (1995). This test was confirmed in the Commission's Discussion Paper in 2005, ¶¶ 259-260. See also Infolab/Ricoh, (1) COMPETITION POL'Y NEWSLETTER (1999). The Commission's Competition Policy Newsletter 1999 No. 1, (February 1999) at pp. 35-37 summarised the conclusion in the Info-Lab/Ricoh case by saying that dominance on the after-market is excluded:

If it is shown that a customer (i) can make an informed choice including lifecycle pricing, that he (ii) is likely to make such an informed choice accordingly, and that (iii) in the case of an apparent policy of exploitation being pursued in one specific aftermarket, a **sufficient** number of customers would adapt their purchasing behaviour at the level of the primary market (iv) within a reasonable time" (emphasis supplied).

Furthermore, information on lifecycle pricing may be available not only from manufacturers of primary and secondary products themselves, but also may be provided by, among other sources, third parties (competitors, specialized reviews and websites, consumer associations reports, etc.) or even by the buyers' ability to spread the information over a series of purchases and by repeat buyers. For example, a majority of printer users are repetitive buyers who already owned a printer in the past, so they are well informed about pricing patterns and lifecycle costs. With the possible exception of *Digital*, in which no formal market definition or finding of dominance and/or abusive behaviour was made, no case of insufficient information has been identified by the Commission.

- 9 Guidelines on Vertical Restraints, OJ No. C-130/1, ¶91 (May 19,2010).
- 10 Chevalier, *Dominance sur un marché des produits secondaires*, EUROPEAN COMMISSION COMPETITION NEWSLETTER (February 1998). In addition to the factors listed in the text, he mentions that if the choice of the primary product is not based primarily on cost (e.g. is based on technical characteristics or reputation) then an increase in the price of secondary products is less likely to affect the choice. New and existing buyers may be individuals or "professionals" who are both likely to be better informed and to be better able to assess all the costs involved. The existence of a second-hand market for the primary products is also relevant. Chevalier also points out that if the secondary market consists of services, the manufacturer supplying them may be able to differentiate between customers who carefully calculate cost and those who do not.
- 11 To facilitate comparisons across printers (and related supplies) and make easy a cost-of-ownership assessment by non-sophisticated customers purchasing printers for home-use, the UK Office of Fair Trading report, *Consumer IT Goods and Services* (2002) led to the adoption of a standard ISO/IEC 27411 - December 2006. This supplied information on inkjet printer cartridge page yields, allowing buyers to compare print yields (average number of printable pages) when using different manufacturers' equipment and to assess cost-per-page when buying a printer. This is important because the main reason why buyers of equipment may not be able to estimate the cost of the aftermarket products over the life of the equipment is lack of convenient comparable information. Similar international standards have been developed for information on monochrome toner cartridges (ISO/IEC 19752 - June 2004) and for color toner cartridges (ISO/IEC 1979 - December 2006) page yield measurements. Regulatory measures to facilitate comparison of lifetime costs seem better solutions than using competition law. In practice, a careful or sophisticated buyer of primary products can estimate what the cost of consumables (and to a lesser extent, repairs and maintenance) will be for the primary product that he is choosing.
- 12 This important point was made by the Court in the Swiss Watchmakers' case, Case T-427/08, CEHR v. Commission, Judgment dated December 15, 2010, (discussed *infra*) ¶ 80.

- 13 See Privy Council decision *Kaisha v. Green Cartridge Company (Hong Kong) Limited (Hong Kong)* [1997] UKPC 19 (30th April, 1997):

18. Furthermore, the ability to control the aftermarket and price the machines on the assumption that the purchasers will buy one's cartridges may actually enhance competition and provide greater choice to consumers, because it will enable manufacturers to compete not only on quality and price but also on the way they divide the cost of their products between the initial outlay and the aftermarket. For example, as Rogers J. pointed out, expenditure in the aftermarket may be treated by the tax authorities as revenue costs and more fully deductible than the capital cost of the machine. Thus a manufacturer who prices the machines lower and the cartridges higher may secure a competitive advantage as against a rival who charges the same lifetime cost in different proportions.

- 14 See OFT Market Definition (2004) paper, ¶ 6.7:

A supplier might not wish to increase prices of its secondary product for existing customers if that would earn it a reputation for exploitation and significantly reduce its ability to attract new or repeat customers to its primary product. Reputation is more likely to be important where suppliers have the prospect of relatively large numbers of new or repeat customers and where undertakings cannot price discriminate between new or repeat customers and other customers.

- 15 For an analysis of when installed base opportunism might occur, see Carl Shapiro & David J. Teece, *Systems competition and aftermarket: an economic analysis of Kodak*, 39(1) ANTI-TRUST BULL., at 135-162, (Spring 1994); see, also, Shapiro, *supra* note 1 at 483-511.
- 16 H. Hovenkamp, *Post-Chicago Antitrust: A Review and Critique*, COLUMBIA BUS. L. REV. 257 - 337 at 289 (2006).
- 17 Case 22/78, [1979] ECR 1869. The Court said it was necessary to decide whether the supply of spare parts constitutes a specific market or forms part of a wider market, and considered only alternative sources of supply of Hugin-compatible parts, and not the wider market for cash registers.
- 18 Case C-53/92 P, [1994] ECR I-667.
- 19 Case C-333/94 P, [1996] ECR I-5951.
- 20 *Supra* note 8, at 41-42.
- 21 Dolmans & Pickering, *The 1997 Digital Undertaking*, EUR. COMPETITION L. REV. 108-115 (1998).
- 22 Chevalier, *supra* note 10. A company considering buying a Digital system could find out how much it would cost; the Commission's point was that this could not be compared easily with the cost of competing systems.
- 23 Commission, COMPETITION POLICY NEWSLETTER, 35 (1999).
- 24 European Federation of Ink Manufacturers (EFIM), Commission Decision COMP/C-3/39.391 EFIM.
- 25 U.K. Office of Fair Trading, Decision dated 20 July 2001, Case CA98/6/2001, ICL/Synstar.
- 26 Commission, XXVI ANNUAL REPORT ON COMPETITION POLICY 37, Novo Nordisk, (1996).
- 27 In *Pelikan/Kyocera*, Pelikan centered some of its (unsuccessful) allegations on the limits placed by Kyocera on its printer warranty if damage was caused to the printer through the use of non-Kyocera toner car-

tridges. Note that this is distinct from the Novo Nordisk case, where warranty of the primary product was disclaimed for any failure either attributable or not to use of compatible secondary products.

- 28 Case T-427/08, CEHR v. Commission, Judgment dated December 15, 2010.
- 29 DG Competition Discussion Paper on the application of Article 82 of the Treaty to exclusionary abuses, ¶¶243-265 (December 2005).
- 30 Communication from the Commission—Guidance on the Commission’s enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings, OJ No. C-45/7, (February 24, 2009).
- 31 *Id.* ¶246.
- 32 *Id.* ¶264.
- 33 O’DONOGHUE & PADILLA, *THE LAW AND ECONOMICS OF ARTICLE 82 EC*, pp. 508-509 (2006).
- 34 Hilti, OJ No. L-65/19, 1988.
- 35 Tetra Pak II, OJ No. L-72/1, 1992; Case C-333/94 P, [1996] ECR I-5951.
- 36 Case T-201/04, Microsoft [2007] ECR II-3601.
- 37 Kühn, Stillman, & Cafarra, *Economic Theories of Bundling and Their Policy Implications in Abuse Cases: An Assessment in Light of the Microsoft case*, (1) EUR. COMPETITION J., 85-121, 103-119 (2005), who say that the efficiency effects of bundling are difficult to evaluate.
- 38 The joint comments of the American Bar Association’s Section on Antitrust Law and Section on International Law on the Discussion Paper conclude that “This fundamental insight regarding the key relationship between the primary market and any related aftermarkets means that a separate examination of a single brand aftermarket under Article 82 is seldom, if ever, appropriate” (p. 35).
- 39 Manufacturers may be able to differentiate in price between new and existing buyers of Services such as maintenance, because services cannot be re-sold.
- 40 Shapiro, *supra* note 1 at 502-504.
- 41 504 U.S. 451 (1992).
- 42 For example, in *ID Security Systems Canada v. Checkpoint Systems*, 249 F.Supp.2d. 622 (E.D.Pa.2003) the Court held that although switching costs were substantial, information was readily available and purchasers were well-informed, so there was no antitrust infringement.
- 43 *Digital Equipment Corp. v. Uniq Digital Technologies, Inc.*, 73 F.3d. 756 (7th Cir.1996).
- 44 In *Queen City Pizza, Inc. v. Domino’s Pizza, Inc.*, 124 F.3d 430 (3d Cir.1997), there was no antitrust infringement because the buyers had agreed to buy pizza dough from Domino’s. The lock-in was contractual.
- 45 *SMS Systems Maintenance Services, Inc. v. Digital Equipment Corp.*, 188 F.3d. 11 (1st Cir. 1999).
- 46 Hovenkamp, *supra* note 16; D. Carlton, *A General Analysis of Exclusionary Conduct and Refusal to Deal - Why Aspen and Kodak are Misguided*, 69 ANTITRUST L.J. 659 (2001); Carlton & Waldman,

Competition, Monopoly and Aftermarkets, Johnson School Research Paper Series #10-06 (2006) argue that competitive aftermarkets are not necessarily efficient and that “there is little that antitrust intervention can do to improve matters, but there is a lot such intervention can do to make matters worse.”

- 47 Goldfine & Vorassi, *The Fall of the Kodak Aftermarket Doctrine: Dying a Slow Death in the Lower Courts*, 72(1) ANTITRUST L.J., 209-231 (2004).
- 48 Guidance on the Commission’s enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings, O.J.N° C. 45/7, February 24, 2009 says, “in the special case of tying in the after-markets, the condition is that the undertaking is dominant in the tying market and/or the tied after-market.” (¶ 50).
- 49 *Id.* ¶¶ 47-62, at ¶ 50.
- 50 Case T-201/04, *Microsoft* [2007] ECR II-3061.
- 51 *Id.* ¶50, footnote 3.
- 52 See Centre for European Policy Studies Task Force Report, *Treatment of Exclusionary Abuses under Article 82 EC Treaty*, 58-62 (2009).
- 53 OFT, *supra* note 2, 414, ¶ 84.
- 54 There would be free-riding competition if one company incurs the R&D costs of the primary equipment but is prevented from recovering those costs by means of profitable sales of the secondary products. This would discourage development of primary products. This is acknowledged by the Commission, as the Guidance Paper says in the section dealing with refusal to supply: “Also, competitors may be tempted to free ride on investments made by the dominant undertaking instead of investing themselves. Neither of these consequences would, in the long run, be in the interest of consumers.” The same comment would apply in the context of tying as in the context of refusal to supply.
- 55 Case C-333/94 P, [1996] ECR I-5951.
- 56 The difficulty with any argument on these lines is that the competition authority is required to foresee the future with more confidence than is usually possible.
- 57 See, generally, Temple Lang, *European competition law and compulsory licensing of Intellectual Property rights—A comprehensive principle*, 4 EUROPARÄTSLIG TIDSKRIFT, pp. 558-588 (2004); Temple Lang, *Anticompetitive non-pricing abuses under European and national antitrust law*, 2003 Fordham Corporate Law Institute, (Hawk ed.) 235-340 (2004) (hereinafter “Temple Lang, Anticompetitive”).

The principal EU cases on refusal to contract, the effects of which are summarized briefly, in chronological order are:

- Case 238/87, *Volvo v. Veng*, [1988] ECR 6211.
- Joined Cases C-241/91 P and C-242/91 P, *RTE and ITP*, [1995] ECR I-743.
- Case C-7/97, *Bronner* [1998] ECR I-7791.
- Case C-418/01, *IMS Health* [2004] ECR I-5039.
- Case T-201/04, *Microsoft* [2007] ECR II-3601.
- Case C-468/06 to C-478/06, *GlaxoSmithKline*, Sept 16 2008.
- Case T-301/04, *Clearstream*, Sept 9 2009.

- Case C-52/09, *TeliaSonera* (not yet decided, a margin squeeze case that has been correctly dealt by Advocate General in his Opinion delivered on 2 September 2010 similarly to a refusal to supply).
- 58 In connection with the *Info-Lab/Ricoh* case, see (1) COMMISSION'S COMPETITION POLICY NEWSLETTER, ¶¶75-90 (February 1999), the Commission official's article on that case, which says, "even if Ricoh had a dominant position it is doubtful whether this would be sufficient to justify imposing an obligation on Ricoh to sell empty cartridges to Info-Lab."
- 59 This point is fundamental, and may seem obvious, but it often seems to be forgotten. It has been frequently repeated, in intellectual property cases, by the Courts, that a refusal to licence is illegal only if it is linked to some other abusive conduct: Case 238/87, *Volvo v. Veng*, [1988] ECR 6211; Cases C-241/91 P and C-242/91 P, *RTE and ITP*, [1995] ECR I-743; Case C-418/01, *IMS Health v. NDC*, [2004] ECR I-5039; Case T-201/04, *Microsoft* [2007] ECR II-3601 ¶¶ 319 ff, 643 ff. But in some cases (not aftermarket cases) the Commission has come close to saying that, even without proof of any specific abuse, a dominant company has a duty to supply so as to create competition downstream. That would mean that the mere absence of effective competition, or of any competition, would be equivalent to an abuse, even if the situation was not due to any conduct of the dominant company. That would make dominance, and not merely abuse of dominance, illegal. This is a regulatory theory, which has no place in competition law. In its interim measures decision in *IMS Health*, the Commission based an order to contract on the ground that, without a contract, there would be a monopoly, because pharmaceutical companies preferred the copyright maps used by *IMS Health*: OJ No. L-59/18, Feb. 28 2002, withdrawn OJ No. L-268/69, 2003; Case T-184/01 R, [2001] ECR II-2349 and 3193. In our view, this decision was clearly wrong, because it suggested that merely because there was a monopoly, even if due only to customer preferences, there was an abuse. H. Hovenkamp, *Unilateral Refusals to Deal, Vertical Integration, and the Essential Facility Doctrine*, University of Iowa Legal Studies Research paper 08-31 (2008) points out that consumers are no better off when a monopoly is shared.
- 60 Article 102(b) expressly requires prejudice to consumers as a condition of an abuse under that clause. The Commission's Guidance Paper, *supra* note 2, describes anticompetitive foreclosure in terms of harm to consumers (¶¶ 19-22) and indicates that foreclosure is legal (and not anticompetitive) if it causes no consumer harm. Under Article 101 conduct "reducing the welfare of the final consumer" is prohibited: Case T-168/01, *GlaxoSmithKline v. Commission*, [2006] ECR II-2969 ¶ 118; but see *Joined Cases C-501/06 P and others, GlaxoSmithKline*, [2009] ECR I-____ October 6, 2009 ¶ 63. The reasons for saying that harm to consumers must be required in all cases of abuse are set out in Temple Lang, *Anticompetitive, supra* note 57, at 250-253.
- 61 If there were competing suppliers of spare parts, a company could be dominant in the primary market but not in the secondary market. However, dominance on the downstream market is not necessary for abuse in margin squeeze cases, see Case C-52/09, *Konkurrensverket v. TeliaSonera* [2011] ECR I- (Feb. 17, 2011).
- 62 The previous sale might not be "equivalent" under Art. 102(c), so there might be no duty to contract under the principle prohibiting discrimination. But the previous contract would show that a sale might be rational, and that there was a "market" for the input. For example, the fact that some buyers buy single-use cartridges rather than refillable ones cannot be a ground of a complaint under Article 102 TFEU. Buyer preferences cannot create an essential facility or make an otherwise lawful practice unlawful. An essential facility, if one exists, is due to the unavoidable needs of competitors, not to the fact that buyers prefer the dominant company's products. This is important, in particular, because some companies have programs under which e.g., used consumables such as empty photocopying and printing cartridges are returned for recycling, motivated by cost savings and environmental protection purposes. The incidental effect is to deprive competitors of the ability to re-use some empty cartridges that are recycled (although consumers may have the choice of buying new or re-used cartridges). These take-back programs may have a major brand-image effect for OEMs, showing their effort to protect the environment; some buyers ask for take-back programs because of environmental concerns. Furthermore, it should be noted that "re-use" does not necessarily mean better protection of the environment, because a refilled consumable cannot be reused indefinitely, and may go to the

land-fill unless refillers have a recycling program of their own for re-used consumables, which seems not to be the norm.

- 63 There is no duty to contract merely to make the competitor's product a *better* product, but there may be a duty to contract if the competitor would otherwise be unable to compete, or if it would be "unreasonably difficult" for it to do so, and if all the other conditions are all fulfilled. The IMS Health judgement (2004 ECR I-5039) ¶ 28 (and the *Bronner* judgment, referred to below in the text) make it clear that there is no duty to contract if there are "alternative solutions, even if they are less advantageous." Even before the Microsoft judgment the Court of Appeal in *Intel Corporation v. Via Technologies* [2002] EWCA Civ. 195, said (¶ 48):

... Magill and IMS indicate the circumstances which the Court of Justice and the President of the Court of First Instance respectively regarded as exceptional in the case before them. It does not follow that other circumstances in other cases will not be regarded as exceptional ... there could be a breach of Article 82 without the exclusion of a wholly new product or all competition. This approach seems to me to be warranted by the width of the descriptions of abuse contained in Article 82 itself. I would, in any event, reject the submission of Counsel for Intel that the IMS test requires the exclusion of all competition from all sources. This was not a requirement in *Oscar Bronner* which referred ... only to all competition from the person requesting the service. Accordingly to the Summary in *IMS* ... must be read in that light. Were it otherwise liability under Article 82 could be simply avoided by a grant of a licence to an unenergetic rival.

- 64 "The circumstance relating to the appearance of a new product ... cannot be the only parameter which determines whether a refusal to licence an intellectual property right is capable of causing prejudice to consumers within the meaning of Article 82(b). As that provision states, such prejudice may arise when there is a limitation not only of production or markets, but also of technical development": [2007] ECR II-3601 at ¶ 647. If the supposed abuse is preventing the claimant from producing a new kind of product (not merely an additional product of the same kind) for which there is a clear and unsatisfied demand, it must be a defense if the dominant company can prove that it already has a business plan to produce the new kind of product that is in question.
- 65 The case law has made it clear that Art. 102(b) applies to limiting the production, marketing, or technical development of competitors, and not merely to limiting the dominant company's own activities. *Joined Cases 40/73 and others, Sugar Cartel—SZV*, [1975] ECR 1663, ¶¶ 399, 482-83, 523-527 ("the system complained of was likely to limit markets to the prejudice of consumers within the measure of Article [82](b) because it gave other producers ... no chance or restricted their opportunities of competing with sugar sold by SZV": ¶. 526); *Case 41/83 Italy v. Commission (British Telecommunications)*, [1985] ECR 873; *Case 311/84, Telemarketing CBEM*, [1985] ECR 3261, ¶. 26; *Case 53/87, CICR v. Renault*, [1988] ECR 6039; *Case 238/87, Volvo v. Veng*, [1988] 6211; *Joined Cases C-241/91P, RTE and ITP ("Magill")*, [1995] ECR I-743 at ¶ 54 ("The applicants' refusal to provide basic information by relying on national copyright provisions thus prevented the appearance of a new product, a comprehensive weekly guide to television programmes, which the applicants did not offer and for which there was a potential consumer demand. Such refusal constitutes an abuse under heading (b) of the second paragraph of Article [82] of the Treaty."); *Case C-41/90, Höfner and Elser*, [1991] ECR I-1979 at 2017-2018 ("Pursuant to Article [82](b), such an abuse may in particular consist in limiting the provision of a service, to the prejudice of those seeking to avail of it": ¶. 30; *Case C-55/96, Job Centre*, [1997] ECR I-7119 at 7149-7150; *Case C-258/98 Carra*, [2000] ECR I-4217; *Case T-201/04, Microsoft*, [2007] ECR I-3601 ¶. 643-648 ("The circumstance relating to the appearance of a new product, as envisaged in *Magill* and *IMS Health* ... cannot be the only parameter which determines whether a refusal to licence an intellectual property right is capable of causing prejudice to consumers within the meaning of Article 82(b) EC. As that provision states, such prejudice may also arise where there is a limitation not only of production or markets, but also of technical development": ¶ 647). *BELLAMY & CHILD, EUROPEAN COMMUNITY LAW OF COMPETITION*, 6th ed., pp. 1025-1026 (2008); *Commission Decision, P&I Clubs*, OJ No. L-125/12, May 19, 1999, ¶¶ 128-133. See E. Elhaage, *Defining Better Monopolisation Standards*, 56 *STANFORD L. REV.* 253 (2003); Temple Lang, *Anticompetitive, supra* note 57 at 235-340; Temple Lang, *The Requirements for a Commission Notice on the Concept of Abuse*

under Article 82 EC, FINNISH COMPETITION LAW YEARBOOK 271-306 (2007); TEMPLE LANG, CENTRE FOR EUROPEAN POLICY STUDIES SPECIAL REPORT (2008); O'Donoghue & Padilla, *supra* note 33, Ch. 4; O'Donoghue, *Verbalizing a general test for exclusionary conduct under Article 82 EC*, in EHLERMANN & MARQUIS, EUROPEAN COMPETITION LAW ANNUAL 2007: A REFORMED APPROACH TO ARTICLE 82 (2008); Vickers, *Abuse of Market Power*, 115 (6) ECON. J., 244 (2005) (who discusses three tests, on sacrifice, as-efficient-competitors, and consumer welfare). Other tests suggested have included a no-economic-sense test and a consumer surplus test).

- 66 The IMS Health interim measures decision of the Commission involved a single market, and would have allowed competitors to use IMS Health's principal competitive advantage to produce identical products.
- 67 Case C-7/97, Bronner, [1997] ECR I-7791, ¶¶ 43-44.
- 68 *Id.* ¶¶ 41-46.
- 69 Temple Lang, *Anticompetitive*, *supra* note 57, at pp. 284-308; Temple Lang, *European Competition Law and Compulsory Licensing of Intellectual Property Rights—A comprehensive principle*, 4 EUROPARÄTTSLIG TIDSKRIFT, 558-588 (2004). Since the mere refusal to give access to any other property or asset cannot be an abuse, it is not clear whether the rules on intellectual property rights are different from those for other kinds of property. However, the Court of Justice has clearly stated several times that refusal to licence an intellectual property right, in itself, cannot be an abuse, and that some "additional element" (which apparently must be a separate abuse) must be present. The Guidance Paper, *supra* note 2 at ¶87 says that consumers may be harmed if the refusal to contract is likely to stifle follow-on innovation. This statement is not limited by reference to any circumstances such as those in the *Microsoft* case. In my view, it is both too vague and too broad. There cannot be a duty to contract merely to enable a competitor to copy or improve on or add to a product already made by the dominant company.
- 70 See the opinion of Advocate General Mazak in case C-52/09, *Konkurrensverket v. TeliaSonera*, delivered on September 2, 2010, and the judgment, dated February 17, 2011, which confirmed that under EU law there can be an illegal margin squeeze even if there is no duty under Art. 102 to supply, and that the company does not need to be dominant in the downstream (retail) market: see, also, case C-280/08 P, *Deutsche Telekom* [2010] ECR I- (October 14, 2010).
- 71 For criteria to be used for deciding whether a price is excessive and contrary to Article 102(a), see Temple Lang, *The requirements for a Commission Notice on the concept of abuse under Article 82 EC*, 2007 FINNISH COMPETITION LAW YEARBOOK (Mentula et al. eds.), 271-305, at 280-284 (2008).
- 72 In *Volvo v. Veng* it was suggested that it might be an abuse to stop producing spare parts, and not to allow others to produce them, if the purpose was to force users to buy new products, or to prevent competitors from providing services for which spare parts would be needed.
- 73 Case T-201/04, *Microsoft*, [2007] ECR II-3601.
- 74 *Id.* at ¶81.
- 75 In *National Carbonising v. National Coal Board* (1976) the Commission finally concluded that NCB had no duty to lower its price to NCC to enable NCC to make a profit on domestic coke, since NCC's difficulties were due to its reduced sales of industrial coke, for which NCB was not responsible. NCC had fewer long-term contracts to supply industrial coke, and so its sales declined. In its decision *BBI/Boosey & Hawkes* OJ No.L-286/36, ¶19 (1987), the Commission said that there is under Article 102 (as it now is) "no obligation placed on a dominant producer to subsidise competition to itself."
- 76 Proof of harm to consumers is required by Article 102(b). Ordering a dominant company to supply an input always creates at least some competition in the downstream market in the short term. This is

not a sufficient reason for imposing a duty to contract. The harm to consumers must be something more than the mere absence of competitors, or absence of choice. In cases in which consumers are not being deprived by the refusal of something that is already available to them, there would need to be clear evidence that competitors would provide some specific advantages, innovations, or developments of importance for consumers. Mere claims that they will ultimately offer better products or services are not enough, and it is important that they would not be able to relax their efforts even if they contracted with the dominant company. In other words, there must be not only scope for added value competition, but also sufficient probability that it will occur and be significant.

- 77 This requirement means that both the dominant company and its competitors continue to be under competitive pressure to develop better products, an essential objective of any rule designed to promote dynamic competition.
- 78 Special features of the *Microsoft* interoperability case that individually or cumulatively seem to have been significant were: Network effects; Exceptional extent and duration of dominance; That the benefits relied on could have been obtained without the conduct in question (Judgment, ¶1154); A pattern of exclusionary conduct; Interoperability had been practiced in the industry and by Microsoft itself; There was no risk that interoperability could lead to mere copying of the whole product; and Because of time lag and disadvantages, competitors would always need to do more than merely provide interoperability. Also, there were high market shares in the downstream market, and Microsoft was not capacity-constrained in that market. A high proportion of competitors in that market were affected. Microsoft's refusal was part of an exclusionary strategy: Disclosure would encourage innovation in the whole industry, including Microsoft; and Reduction of innovation harms consumers through reduced choice and lock-in of users. In any refusal to contract case it may be necessary to balance short-term effects of promoting competition against long-term effects of reducing incentives to invest. But the legal rules should be designed so as to avoid, as far as possible, making the test merely a balancing test, which would be contrary to legal certainty. This problem is not the result of the specific phrase used by the Commission. This is distinct from the problem of balancing the short-term pro-competitive and anticompetitive effects of the conduct in question.
- 79 A software company can limit access to its software upgrades / software support tools to customers who licensed the original software, and not make those upgrades available for licensing stand-alone, at least where the software company is not dominant for the original software.
- 80 Commission, *Pharmaceutical Sector Inquiry: Preliminary Report* (November 2008).
- 81 See, generally, Centre for European Policy Studies, *Report of Task Force, Treatment of Exclusionary Abuses under Article 82 of the EC Treaty*, 21-28 (2009).
- 82 Commission, *Pharmaceutical Sector Inquiry: Final Report* (July 2009). See Temple Lang, *Current Competition Law Questions for Pharmaceutical companies*, CURRENT DEVELOPMENTS IN EUROPEAN AND INTERNATIONAL COMPETITION LAW, 16th St. Gallen International Competition Law Forum 2009 (Baudenbracher ed.), 79-111 (2010).
- 83 553 U.S. ____ (2008).
- 84 Commission decision *Decca Navigation System*, OJ L-43/27 (1989). Factual issues might arise if the change involved genuine improvements, but also made competitors' products incompatible with the improved equipment. If the improvement involved only a change in software, there might be an obligation, in order to avoid an abuse, to disclose the new software to competitors.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Antitrust: A Good Deal for All in Times of Globalization and Recession

Bruno Lasserre

Autorité de la concurrence

Antitrust: A Good Deal for All in Times of Globalization and Recession

*By Bruno Lasserre**

With the economic recession taking over from the financial crisis of 2008, attention has focused less on issues such as merger review and State aid control, and more on antitrust properly speaking, meaning the prohibition of cartels and abuses of dominance, as well as the enforcement of this prohibition by means of administrative fines imposed on corporations and/or of criminal penalties directed to individuals. Among other items, this agenda has included the following questions: 1) whether corporate fines are excessive or indeed misdirected and should be replaced in whole or at least in part by individual penalties; 2) whether antitrust enforcement itself is a luxury good or even an idea of yesterday, and should be abandoned or at least significantly relaxed.

In the following pages, I will briefly address these two issues. I will do so in reverse order, since discussing the significance of public antitrust enforcement via administrative fines and/or criminal penalties makes little sense if this enforcement has become irrelevant in the first place. I will not deal with the separate issue of private enforcement.

| *President of the French Autorité de la concurrence

I. Antitrust Enforcement in Times of Recession and Globalization: Putting the Issue in Context

When the world's economy and financial system entered into a time of unprecedented turmoil at the end of 2008, I expressed the view¹ that it was not only important, but indeed vital, that competition enforcers get their act together and meet the challenges raised by this new situation, within the scope of their mission and with their limited means.

The most urgent of those challenges, I recalled, was to answer the short-term concerns raised by a stumbling financial and banking system and by an anticipated slowdown of the real economy, while at the same time keeping an eye fixed on the long-term vision of a competitive economy that delivered merit-based benefits for corporations and consumers alike. Available means to match those challenges, I advocated, included adapting the processes put in place to scrutinize the competitive impact of corporate mergers and public subsidies before they occurred, while not compromising on the need to make such an upfront assessment instead of standing by until some of these actions had actually jeopardized growth and welfare.

Two years later, it seems that merger review and, as far as the European Union is concerned, State aid control have gone down rather well during the downturn. A number of banking mergers were planned at the peak of the financial crisis and in its immediate aftermath, sometimes raising competition concerns, but almost all of them have been allowed to proceed thanks to the commitments negotiated between the parties and the competition authority or authorities in charge of these deals in order to alleviate such concerns. In France for instance, the recent *Banque Populaire/Caisse d'Épargne* case² shed some light on how a banking merger could be cleared thanks to an upfront dialogue between the parties and the enforcement agency on how the crisis affects (or not) the range of remedies available to meet the competitive concerns raised by the deal.

As for State aid cases, the European Commission has been praised for taking a series of swift decisions granting survival packages to distressed financial institutions while insisting, first, that they be devised in the least distortive way; second, that their features be sufficiently consistent so as to preserve a level-playing field throughout the European market; and, third, that the aid granted be monitored and reimbursed as soon as market conditions allowed for the distressed institution's recovery.

While no one knows whether or not this interim diagnosis will remain valid once the time comes to make a full checkup, it is therefore fair to say that, overall, merger review and State aid control seem to have adapted to recent events

TWO YEARS LATER, IT SEEMS THAT MERGER REVIEW AND, AS FAR AS THE EUROPEAN UNION IS CONCERNED, STATE AID CONTROL HAVE GONE DOWN RATHER WELL DURING THE DOWNTURN.

by addressing the individual concerns related to the failure or potential failure of a number of key market players, while taking the steps required to make sure that customers and consumers do not end up paying the price of these operations of market maintenance. As a consequence, the debate in this regard has somewhat receded.

MEANWHILE, THE ATTENTION HAS FOCUSED ON ANTITRUST PROPERLY SPEAKING, MEANING THE PROHIBITION OF CARTELS AND ABUSES OF DOMINANCE, AS WELL AS THE ENFORCEMENT OF THIS PROHIBITION BY MEANS OF ADMINISTRATIVE FINES.

Meanwhile, the attention has focused on antitrust properly speaking, meaning the prohibition of cartels and abuses of dominance, as well as the enforcement of this prohibition by means of administrative fines imposed on corporations and/or of criminal penalties directed to individuals. Among other items, this agenda has included the following questions: 1) whether corporate fines are excessive or indeed misdirected and should be replaced in whole or at least in part by individual penalties; and 2) whether antitrust enforcement itself is a luxury good or even an idea of yesterday, and should be abandoned or at least significantly relaxed.³

In the following pages, I will briefly address these two issues. I will do so in reverse order, since discussing the significance of public antitrust enforcement via administrative fines and/or criminal penalties makes little sense if this enforcement has become irrelevant in the first place. I will not deal with the separate issue of private enforcement.

II. Is Antitrust Enforcement Irrelevant and Should It Be Relaxed or Even Relinquished?

Two issues are generally put forward in support of the idea that antitrust enforcement has become less relevant or even irrelevant today. These issues are often confused or conflated. They may be related to some extent, but in my mind they raise different questions and therefore call for a separate assessment.

The first contention (a) is that antitrust enforcement is a “luxury good,” meaning that, while it may be afforded in good economic times, it must be relaxed in bad economic times. The second argument (b) is that antitrust enforcement is “so yesterday,” meaning that, while it could be accommodated while Western economies flourished, it cannot be tolerated anymore in times of globalization. The first variant therefore supports a temporary relaxation of antitrust enforcement, while the second advocates its permanent dismantling.

A. IS ANTITRUST ENFORCEMENT TEMPORARILY UNAFFORDABLE?

Deciding whether antitrust enforcement is a luxury good or not is a matter of personal belief. I have a personal opinion on the matter, which is certainly sub-

jective and which some could perhaps view as biased, but which I hope all can accept as thoughtful given my background. Rather than advocating my opinion in the following lines, I will rather try to shed some light on a few elements of context to be considered when addressing the issue.

1. Antitrust Law is Intended to Benefit Both Corporations and Consumers

Perhaps one of the most remarkable developments of competition policy in the last decade is the effort put by competition authorities to root their decisions on when, why, and how to enforce the law in economic analysis. An ever-larger number of competition authorities have joined the trend as evidenced by these authorities prioritizing cases that are most likely to damage consumer welfare; evidencing robust theories of harm before challenging mergers, horizontal and vertical agreements, and unilateral conducts; balancing their likely adverse impact on competition with the efficiencies and benefits that proposed combinations or individual strategies are likely to yield; and, when necessary, making sure that remedies imposed on individual firms guarantee or restore competition without unduly chilling their freedom to invest, to innovate, and to compete on their merits.

PERHAPS ONE OF THE MOST
REMARKABLE DEVELOPMENTS OF
COMPETITION POLICY IN THE
LAST DECADE IS THE EFFORT
PUT BY COMPETITION
AUTHORITIES TO ROOT THEIR
DECISIONS ON WHEN, WHY,
AND HOW TO ENFORCE THE LAW
IN ECONOMIC ANALYSIS.

The International Competition Network (“ICN”), which is not an inter-governmental organization such as, for instance, the World Trade Organization (“WTO”), but rather an informal network now bringing together almost 120 competition enforcement agencies in the world, has played a decisive role in bringing about and accompanying this modernization.⁴ On a more personal note, I have been—and I remain—a long-standing and committed advocate of such a shift towards a so-called “more economic and effects-based approach.” I also hope to have driven the former *Conseil de la concurrence*, and now the *Autorité de la concurrence*, down this road since I took the helm of the agency in 2004.⁵

The dedication with which we make sure that competition policy efficiently drives, rather than unduly blocks, intelligent and courageous corporations in their efforts to take risks, to innovate, to market better products or services at better prices, and to deliver them to end-consumers should, however, not make us forget why competition law was created in the first place. Making firms more efficient and more profitable is clearly part of the picture. But competition law, as it has stood in Europe since 1957 or in France since 1986, also incorporates the notion that a fair share of the extra profits yielded by this additional efficiency be passed on to consumers.⁶ In other words, it walks on two legs: driving cor-

porations to do their best is not only in their self-interest, it is also in the general interest of citizens as a whole.

A similar vision, I think, is encapsulated in one sentence of the statement made by then-Senator Obama before the American Antitrust Institute (“AAI”) in the course of the 2008 presidential campaign, which could, I guess, be taken to reflect a bi-partisan view on the role of competition law and policy on the U.S. side of the Atlantic Ocean: “Antitrust is the American way to make capitalism work for consumers.”⁷ It is also this vision, according to which a competitive marketplace delivers growth, jobs, and welfare that benefit society as a whole, that led the French Parliament, with the support of the then-existing *Conseil de la concurrence*, to pass a legislative package revamping our competition enforcement system in 2008/2009, an aim which, it is worth being stressed, earned nearly unanimous support on the benches of the National Assembly and Senate.⁸

2. In Practice, Antitrust Enforcement Actually Balances Corporate and Consumer Welfare

The above is not just a narrative. It is fact-based. A number of competition authorities regularly publish evaluations of the benefits attached to their actions in the field of antitrust enforcement. These evaluations apply rule-of-thumbs hypotheses regarding the likely consequences of cartels and abuses of dominance, notably in terms of price increases, and the correlative benefits of staffing them

THE AUTORITÉ DE LA
CONCURRENCE IS UNDER
A DUTY TO SYSTEMATICALLY
ASSESS, ON A CASE-BY-CASE BASIS,
WHAT ADVERSE CONSEQUENCES
ANTITRUST PRACTICES ARE LIKELY
TO HAVE ON CONSUMERS AS WELL
AS ON THE GENERAL ECONOMY,
AND, CORRESPONDINGLY,
WHAT BENEFITS ITS ENFORCEMENT
YIELDS FOR SOCIETY.

via antitrust enforcement. Such hypotheses can of course be challenged, like all hypotheses. But they are transparent, based on robust economic doctrine, and generally subject to some degree of independent scrutiny by a public body other than the competition authority itself. It is, therefore, difficult to discard them completely.

The figures that these competition authorities put forward are eloquent. I will not detail them for the very simple reason that I am not best-placed to do so since, unlike most of its counterparts, the French *Autorité de la concurrence* has not published such outcome evaluations to date. However, it is under a legal requirement, each time it intends to impose a corporate fine,

to make an assessment of the economic harm caused by the cartel, agreement, or abuse of dominance that is being investigated. In other words, the *Autorité de la concurrence* is under a duty to systematically assess, on a case-by-case basis, what adverse consequences antitrust practices are likely to have on consumers as well as on the general economy, and, correspondingly, what benefits its enforcement yields for society.

The main cartel decision taken by the *Autorité* in 2009, which concerns a case of collusion among the three majors of the temporary work sector in France, illustrates this exercise. The turnover related to the collusion, which was proven to have lasted two years (but may have been around for a longer period), amounted to EUR 5 billion. The *Autorité* assessed the overcharge to be in the magnitude of 0,5 percent, while the parties' data put it between 0 and 1,4 percent.

If one applies the method used by the United Kingdom's Office of Fair Trading ("OFT") for evaluating the outcome of its cartel work, but using the likely overcharge assessed by the *Autorité* (0,5 percent) instead of the standard rule-of-thumb of 10 percent, one can estimate that the *Autorité's* decision saved EUR 138 million. If one applies the methods used by the U.S. Department of Justice's Antitrust Division ("DOJ") and by the European Commission's Directorate General for Competition ("DG COMP") at the time of writing, the figure ranges around EUR 500 million and EUR 2,7 billion, respectively.

One must bear in mind the fact that all these figures focus on the "price effect" of the temporary work cartel (artificial overcharge borne by customers) and leave aside its "volume effect" (temporary jobs foregone due to their increased cost). They also focus on the punitive dimension of the *Autorité's* decision (on the participants to the cartel) and leave aside its deterrent effect (discouraging these same firms as well as other corporations from colluding in the future).

Added to the administrative fines imposed by the *Autorité* in this case (EUR 94 million), the total financial saving for consumers resulting from antitrust enforcement (EUR 232 million), in one single decision, is thus much greater than the agency's annual budget (EUR 20 million).

3. Relaxing Enforcement in Times of Recession Would, In Fact, Mean Letting Consumers Down

This context may be of use when thinking about whether or not antitrust enforcement should be relaxed in bad economic times. The full story reads as follows: When firms collude or abuse their market power, instead of just trying to do their best, they do not simply break the law, they also deprive businesses that operate downstream (either as manufacturers or as distributors), as well as end-consumers, of the tangible welfare benefits of the market economy. Cartels and abuses of dominance are a legal problem (economic crime) in the first place because they are a problem for the economy (inefficient behavior that leads to catastrophic market results, as evidenced by the temporary relaxation of antitrust enforcement in the United States during President Roosevelt's first mandate (1933-1937),⁹ before the administration reverted to trustbusting.¹⁰) But these abuses also are a political problem: How can we expect ordinary citizens to support the market economy if we "relax" while a few corporations rob them from their "fair share" of the profits delivered by the market economy?

Some of us might thus be entitled to claim that antitrust enforcement is not a “luxury good,” but rather a “base product,” and especially so in dire economic times during which it is often noted that cartels and abuses are more frequent. This is why it stands at the heart of Europe’s single market and strategy for growth and welfare in the 21st century, as stressed in Mario Monti’s recent report on the future of the EU¹¹.

B. IS ANTITRUST ENFORCEMENT DEFINITELY UNDESIRABLE?

Alongside the idea of a (supposedly) temporary relaxation of antitrust enforcement, a distinct claim has been made that, whatever the added value of antitrust enforcement, we can definitely not afford it any longer and must resign ourselves to abandon this hallmark of a perhaps generous, but sadly bygone, era. This claim further alleges that this change has been made unavoidable by the evolution that the world has been experiencing over the last decades and that is being, perhaps, accelerated to some extent by the current crisis, namely the advent of a more multi-polar economy. The script of this story-in-the-making is that, until and unless countries that presently have a robust antitrust enforcement regime in place (as well as, for that matter, pieces of legislation aimed at addressing other issues of general interest such as safety regulations or financial regulations) relinquish these unconscionable checks on unfettered markets, countries that currently do not have equivalent instruments in place will be at a decisive advantage in the run for economic growth.

Such a call for an abandonment of antitrust enforcement is not wholly unheard of. History tells us that, in past periods of difficult times, such an abandonment has been advocated in a variety of forums, be it the press, behind closed doors, or in the courtrooms. For instance, the well-known U.S. Supreme Court *Appalachian Coals*¹² case has recently found a distant echo, both in time and in space, in the Paris Court of Appeal *AMD*¹³ case. In this case, the Court found

that a very sophisticated, five-year long, nationwide cartel run by the bulk of steel traders was “serious in theory,” but “largely mitigated in practice” by “a context of economic crisis, both general and sector-specific.”¹⁴

Is this story a true story? If yes, then it is worth being listened to and thought about. After all, it is only reasonable to adjust one’s standard of living to one’s budget.

Answering the question of antitrust’s desirability is made somewhat difficult by the fact that the case for this turnaround on antitrust enforcement is often made with an assertive tone, but rarely elaborated upon in detail. As far as I can see it, though, there are at least two underlying assumptions: “Antitrust kills business,” on the

ANSWERING THE QUESTION OF ANTITRUST’S DESIRABILITY IS MADE SOMEWHAT DIFFICULT BY THE FACT THAT THE CASE FOR THIS TURNAROUND ON ANTITRUST ENFORCEMENT IS OFTEN MADE WITH AN ASSERTIVE TONE, BUT RARELY ELABORATED UPON IN DETAIL.

one hand; “Antitrust kills Europe” (or the United States or indeed any other jurisdiction where it is part of the law), on the other hand.

Are antitrust enforcement regimes indeed millstones that prevent business from being powerful and efficient enough on the global marketplace? I use the terms “powerful” and “efficient” because the basis tenet of competition law is to drive corporations to make efficient use of their market power and to step in only when they are likely to be making, either collectively (cartels and other anticompetitive agreements) or individually (abuses of dominance), an inefficient use of such a power. This story actually has two limbs: that competition law would unduly prevent the formation of market power and/or would also unduly prevent its exercise.

1. Antitrust Law Does Not Stand in the Way of Legitimate Corporate Growth in a Globalized Economy

Frankly, I do not think that competition law enforcement acts as an illegitimate or unnecessary hurdle between corporations and market power. It is extremely rare for corporate plans to merge to be challenged or blocked by competition authorities. The test on which most modern merger regimes in the world rest is not whether the deal is likely to result in the creation of a dominant market position; rather, it is whether it is likely to result in a substantial lessening of competition—something a merger may or may not lead to, depending on the facts of the case, and which may nonetheless be offset by the efficiencies that the merger can bring about for the benefit of consumers, again depending on the facts of the case. The merger landscape abounds with recent cases in which competition authorities all around the world have given the go-ahead, with or without conditions, to mergers eventually giving birth to firms that are fully able to compete on the global or regional scene.

FRANKLY, I DO NOT THINK THAT COMPETITION LAW ENFORCEMENT ACTS AS AN ILLEGITIMATE OR UNNECESSARY HURDLE BETWEEN CORPORATIONS AND MARKET POWER.

To take the European example and to stick to the steel industry, which has undergone successive waves of restructuring and expansion over the last decades, the French number one player, Usinor, in 2001 was allowed to merge with its fellow flag carriers, Arbed from Luxembourg and Aceralia from Spain, thus giving rise to a worldwide player, Arcelor, that later merged with Anglo-Indian Mittal Steel in 2006. The past year provides innumerable examples of other European or American corporations being allowed to combine into global businesses; these examples featured world or regional leaders in their field: *GDF/Suez* in the gas industry; *Kraft Food/Cadbury* in the food sector; *British Airways/Iberia* in the air transport business; *BNP-Paribas/Dexia* in the banking industry; *NBC Universal/Comcast* in the entertainment sector; *News Corp/BSkyB* in the media and communications business; *Schneider Electric/Areva T&D* in the power transmission and distribu-

tion sector; *Oracle/Sun* on the software market... to name but a few of the hundreds of deals that were cleared in the last year by the European Commission. The same goes with the *Autorité*, which has not hesitated since 2009 to clear mergers creating such big players as *Veolia Transport/Transdev* in the transport sector or the already mentioned *Banques Populaires/Caisses d'Épargne* in the banking business.

WE THEREFORE STAND A WORLD
APART FROM THE IDEA
THAT MERGER REVIEW IS AN
OBSTACLE TO BIGNESS.

We therefore stand a world apart from the idea that merger review is an obstacle to big-ness. What merger review does, instead, is screen corporate deals in order to make sure that increases in corporate market power often attached to increases in corporate size do not give rise to situations where a firm enjoying unconstrained market power is in a position to charge higher prices to consumers.

2. Antitrust Enforcement Does Not Prevent Business From Adjusting to the Global Recession

Likewise, I do not see how a convincing case that antitrust enforcement unduly obstructs the conduct of business can be made. Again, almost no anticompetitive practices, be they agreements or unilateral conducts, are prohibited *per se*, most of them being forbidden because of their object or (actual or potential) effects on consumers. The main exception to this statement relates to cartels; these are considered to be “unjustifiable” by the OECD¹⁵ and therefore prohibited *per se* in the United States, as well as deemed anticompetitive in view of their very object by the European Union and by its twenty-seven Member States. I have never heard a serious economist support the idea that this approach to cartels is misplaced. And yet, competition law, as it stands in Europe, does not totally close the door to the justification of a cartel,¹⁶ although it requires that this be done on competitive grounds and on the basis of a case-specific assessment.

The truth is that firms are legally barred from justifying cartels on the mere basis that it is profitable for them to plan their production in an “orderly” fashion (in effect restricting output or allocating markets or customers) or to set prices at a “fair” level (in effect fixing prices) in order to escape the pressure resulting from competition. It is in the nature of such “trusts” and “conspiracies” to be beneficial for their authors or, in any case, to appear as such at the time they are entered into. This is why, the step-stone of antitrust law being that corporations must behave “autonomously” on the marketplace¹⁷ (i.e. defend and promote their interests on their own), they must prove that, whenever they enter into an agreement, this agreement is likely to result in “appreciable objective advantages that compensate the resulting disadvantages for competition”¹⁸ (i.e. to benefit consumers and not only themselves).

The constitutional “policy” behind this legal reasoning, it is worth recalling, is that the aim of European antitrust law is not narrowly defined as “promoting economic efficiency,” but more broadly as “promoting a competitive structure and process on the marketplace.”¹⁹ This statement dates back to the inception of European antitrust law, but it remains valid in full: When the European General Court clearly and willingly raised the subject of whether or not it still held true fifty years on, the European Court of Justice clearly and willingly replied in the affirmative. Economic schools of thought (and especially the “Chicago School”) have therefore significantly informed and, in my mind positively enriched, the political philosophy, legal techniques, and economic reasoning that back antitrust enforcement, but have not led European courts and enforcers to turn around on long-standing constitutional choices, legal precedents, and economic wisdom. As a result, contemporary European antitrust enforcement has kept its original balance, while becoming more sophisticated, which is a good thing.

This explains, in particular, why “crisis cartels” have occasionally been accepted by competition enforcers or competent jurisdictions, but never so on the basis of a general crisis or even of a sector-specific crisis. In other words, “crisis cartels” do not enjoy a specific treatment and are only open to individual justification to the extent that they meet the standard conditions contemplated by the law (both at European and national levels) for justifying agreements that would be prohibited absent such conditions. It is therefore necessary that: 1) they “contribute to improving the production or distribution of goods or to promoting technical or economic progress” (efficiency); 2) “while allowing consumers a fair share of the resulting benefit” (fairness); 3) “not going beyond what is indispensable to attain these objectives” (necessity); and 4) “not being liable to eliminate competition” (proportionality). And the fact is that, given the magnitude of the harm that they can cause to competition and consumers, it should be correspondingly more difficult to show that they are likely to produce efficiencies that offset this harm than in the case of other anti-competitive practices. The law precisely contemplates that, the more business practices are liable to hurt consumers, the more these practices must be shown to produce positive effects that will benefit consumers.

So antitrust enforcement not only is not intended to hurt business but, in fact, does not lead to such outcomes. What it does, however, is make sure that when firms need to adjust to changing economic circumstances, including by contemplating a “crisis cartel,” they make a convincing case, resting on objective and robust evidence, that they are not adjusting at the expenses of consumers and citizens.

What abandoning antitrust enforcement thus means, in effect, is losing sight of this very neat balance.²⁰

CONTEMPORARY EUROPEAN
ANTITRUST ENFORCEMENT HAS
KEPT ITS ORIGINAL BALANCE,
WHILE BECOMING MORE
SOPHISTICATED, WHICH
IS A GOOD THING.

3. The World Economy Needs a Level Playing (Competition and Antitrust) Field

The second main assumption behind the “antitrust abandonment music” played by modern-days Hamelin pipers is that Europe (or the United States, Japan, and so on) have no chance of surviving in today’s world if they remain at such a regulatory disadvantage with other parts of the world that do not have such a regime in place. I will make two related points on this argument which, as paradoxical as it may be, captures a very important truth about the central role of competition and policy in political systems based on the rule of law and on market economy.

THE STORY OF COMPETITION
LAW OVER THE LAST
CENTURY IS THAT OF AN
EVER-LARGER DISSEMINATION.

tion and policy in political systems based on the rule of law and on market economy.

First, things change. The story of competition law over the last century is that of an ever-larger dissemination. The ICN started as a pioneer group of 15 or so jurisdictions. Its membership today extends to more than 100 jurisdictions. It initially focused on how to make merger review more consistent and efficient, in order to avoid jeopardizing pro-competitive corporate deals because of lack of coordination, undue delay, red tape, or flawed economic analysis, although its mandate was, of course, broader. It now routinely enables competition enforcers from around the world to: share insight and experience on policy, substance, and processes; learn from one another; help each other; converge voluntarily on best practices (or occasionally to understand why they differ); and inform and, if appropriate or required, coordinate on ongoing cases. This results in a more efficient enforcement that benefits not only competition agencies themselves, but ultimately also businesses and consumers who benefit from a level-playing field and from consistent outcomes.

Second, this cooperative trend is neither a miracle of nature nor a given. Its development has been derived from the fruits of dedication, persuasion, and emulation. Its future continuation will require constant effort and care. At the same time, it is not sufficient in itself. We have to be realistic and stay aware that simply having a competition law regime in place on paper is of limited use if it is not effectively implemented, monitored, and advocated, as well as sometimes protected.

But what strikes me is that, in recent months, competition enforcers have not been alone in doing the job. To take the French example, both law-makers and Government executives have repeatedly expressed the view that what we needed was to go forward,²¹ not backward. Europe, as well as the United States, have benefited enormously from the rise in international trade that has been made possible not only by the lowering of public tariffs, barriers and subsidies, but also by the curb put on private obstacles to free and fair commerce, including cartels, bid-rigging, and abuses of dominance. Globalization means that these rules should be shared by all of those who play the game of international trade, and not that they should be dismantled where they exist. What the world needs is a truly level playing field, based on rules that are shared by all and implemented

on a reciprocal basis—not the increased market fragmentation and cartelization that have historically proven to damage consumer welfare, to slow economic recovery, and to induce trade wars such as those in the run-up to World War II.

In a recent book, David Gerber has brilliantly analyzed the twice-aborted plans to establish a global framework for competition enforcement,²² first with the Havana Charter in the 1940's and second with the WTO in the 1990's. The fact that these efforts have failed to date does not prevent us, in my view, from thinking about other ways of moving forward, both by fostering voluntary convergence in multilateral forums such as the OECD and by pushing for reciprocity in bilateral trade agreements. Neither should it deter us from dwelling on the good work done by the ICN over the last decade.

Merger review, antitrust enforcement, and competition advocacy should be an integral part of this global commitment, of course, but so should openness and non-discrimination of public procurement and public tender, on the one hand, and the control of public subsidies, on the other hand. Subsidies, either directly or indirectly granted on public funding, have the potential of creating substantial bias on the marketplace, not least when the firms that have benefited from them use them to develop an overseas market, to acquire foreign assets, or to bid for public contracts in other jurisdictions. Existing disciplines, notably those entered into within the framework of the WTO,²³ should thus be strictly enforced. The European Union, which has developed state-of-the-art experience in that regard with its own rules on State aid,²⁴ could usefully make it available to other jurisdictions.

I FIND IT HARD TO SUSTAIN THE
IDEA THAT ANTITRUST LAW AND
POLICY WEAKEN THOSE
COUNTRIES THAT ENFORCE THEM.

In conclusion, I find it hard to sustain the idea that antitrust law and policy weaken those countries that enforce them. Rather, they provide a strong reason for convincing our trading partners that the benefits associated with international trade imply that we mutually enforce rules prohibiting both undue public obstacles and private impediments to interstate commerce.

This leads us to the issue of how these standards can be best enforced.

III. Is Antitrust Enforcement via Corporate Fines Misconceived and Should It Be Phased Out in Part or in Full to the Benefit of Individual Penalties?

As with the question addressed in the previous section of this paper, the ongoing discussion on the relevance of corporate fines broadly rests on two claims that are often intertwined but that call, in my view, for a separate look. The first claim

(a) is the notion that corporate fines have become excessive in level and should be seriously lowered. The second one (b) is the case that labels corporate fines as misguided and claims that they should be replaced by, or at least mixed with, other tools, namely individual penalties such as debarment (also termed disqualification) or indeed jail terms.

These claims differ in content, but not necessarily in outcome, since both of them could eventually result in a phasing out, in part or in full, of antitrust fines.

A. HAVE CORPORATE FINES BECOME EXCESSIVE IN AMOUNT AND SHOULD THEY BE LOWERED?

European (public) enforcement of antitrust rules has relied, since its inception in the 1950's and 1960's, on administrative fines imposed on guilty corporations by specialized authorities, acting under the control of review courts.

This is not to say that criminal penalties are not available in parallel, in order to sanction individuals who are found guilty of committing an antitrust offence. This criminal track is historically absent at European level, since the European Union lacked competence in the criminal field.²⁵ However, it exists at the level of Member States, where it varies in both form and intensity. In France for instance, criminal penalties were historically the main instrument available for enforcing antitrust rules. Antitrust criminal law was nevertheless seldom implemented, as public prosecutors did not prioritize it and criminal judges were reluctant to enforce it.

In other words, antitrust criminal policy was a failure; a situation which eventually led the Government to set up an independent public authority specialized in enforcing antitrust rules via administrative fines imposed on guilty corporations which, pursuant to a series of reforms, ultimately became the *Autorité de la concurrence*. The Code of Commerce still provides for criminal penalties going up to four years of imprisonment against individuals,²⁶ but, to date, this provision has been rarely applied. I will come back to current prospects in that regard in the following section of this article.

So the distinct characteristic of the European Union and of those of its 27 Member States that have (as is the case of most of them) modeled their antitrust enforcement regime on the one in place at the European level is the central role of corporate fines to, first, punish firms found guilty of participation in a cartel or another anticompetitive agreement or abuse of dominance, and, second, deter them, as well as other corporations, from committing such infringements.

1. Corporate Fines Have Increased in Order to Become More Deterrent

It is trite to say that the overall amount of those fines has significantly increased over the recent years, both at European level and in a number of Member States.

The phenomenon is not new. Actually, the leading European precedent on how to set antitrust fines,²⁷ that still today grounds and governs much of the current case-law of the European General Court and Court of Justice, was born out of a fully transparent decision taken by the European Commission in the early 1980's to increase the general level of antitrust fines as compared to those achieved under its previous policy. According to public records, this decision was based on the fact that, more than twenty-five years after the birth of the European Union (then called the European Communities), serious business malpractice, in particular market-sharing, output-restricting, customer-allocating, and price-fixing agreements, had not visibly diminished either in number or in intensity.

This change in policy, which was endorsed by the European judicature, triggered a trend of increasing fines that was continuing when the European Commission published its second fining guidelines in 2006.²⁸ In this context, the first generation of guidelines, published in 1998²⁹ after the European General Court (then the European Court of First Instance) had invited the European Commission to do so, was intended first and foremost to make the Commission's fining practice more transparent and more consistent, by making known in advance what criteria the Commission used on a case-by-case basis and in which way it did so. The second set of guidelines had the distinct objective of increasing yet further the general level of antitrust fines. This was fully acknowledged by then-Commissioner Kroes, whose famous message to companies contemplating a violation of European rules outlawing cartels and other anticompetitive practices was clearly set on deterrence: "Don't break antitrust rules; if you do, stop as quickly as possible; once you've stopped, don't do it again."³⁰

This policy produced well-known results. Total corporate antitrust fines³¹ increased from EUR 540 million in 1990/1994 and EUR 293 million in 1994/1999 to EUR 3.463 million in 1999/2004 and EUR 9.761 million in 2004/2009. This trend is not unequivocal. For instance, a closer look at the last five years reveals that yearly fines, that amounted to EUR 1.846 million in 2006 and peaked to EUR 3.338 million in 2007, receded to EUR 2.270 million in 2008 and to EUR 1.623 million in 2009, before reaching EUR 3.057 million in 2010. However, if one accounts for the facts that the total amount is dependent notably first on the nature of the offences adjudicated each year by the European Commission (which has been consistently focusing over the recent period on hardcore, often international or European-wide, cartels); second, on yearly output (which varies to a significant extent if one looks, not at the number of cases handed out, which has consistently ranged between 6 and 8 since 2006, but at the number of individual firms involved in those cases, which evolved between 37 in 2008 and 69 in 2010); and third, on the individual situation of these firms

IT IS TRITE TO SAY THAT THE OVERALL AMOUNT OF FINES HAS SIGNIFICANTLY INCREASED OVER THE RECENT YEARS, BOTH AT EUROPEAN LEVEL AND IN A NUMBER OF MEMBER STATES.

(power on the affected market, overall power, duration of participation, specific role, etc.), it is fair to say that fines have increased to a very significant extent during the last decade.

In France, the increase in fines was triggered in the first place by a radical overhaul of the sentencing provisions of the Code of Commerce that occurred in 2001.³² The two main changes decided by the Government and the Parliament consisted, on the one hand, in increasing the legal maximum from 5 percent of the French turnover of the firm liable for the infringement to 10 percent of the worldwide turnover of the group to which it belonged, and, on the other hand, in providing that fines should not only be proportionate to the “seriousness of the infringement,” to the “importance of the harm caused to the economy” and to “the individual situation of the firm or of the group to which it belongs,” but should also incorporate, where applicable, a separate premium in case of “reiteration.”³³

The preamble to the bill makes it clear that the intent of Parliament was to increase the severity of the fining regime in order to meeting four challenges: first, matching a trend of ongoing recidivism among law-breakers; second, accounting for the increasing globalization of business strategies, including (but not limited to) anticompetitive practices; third, putting an end to the frequent circumvention of the previous rules by way of artificial spin-offs of daughter companies; and fourth, making domestic fining rules consistent with the standard existing at the level of the European Union.³⁴

Since then, the overall level of fines imposed by the *Conseil de la concurrence* and now by the *Autorité de la concurrence* has significantly increased, in a way that can nevertheless not be compared to what occurred at European level given the differences of scope and nature between the infringement cases adjudicated by the European Commission (mainly international or European-wide cartels and large abuses of dominance) and those handled by the *Autorité* (that include not only nationwide cartels and abuses of dominance, but also a variety of regional or even local anticompetitive agreements, bid-rigging, vertical restraints and unilateral conducts).

While total figures ranged in the vicinity of EUR 60 to 65 million per annum in the first half of the 2000's—when cases handled by the agency were still being fined in accordance with the legal standard applicable until 2001—in compliance with the principle of non-retroactivity, they have subsequently amounted³⁵ to EUR 754 million in 2005,³⁶ EUR 128 million in 2006, EUR 221 million in 2007, €631 million in 2008,³⁷ EUR 206 million in 2009, and EUR 442 million in 2010.³⁸ Over the same period, the number of fining decisions has diminished (31 in 2005, 13 in 2006, 24 in 2007, 16 in 2008, 15 in 2009, and 12 in 2010), as well as the total number of firms fined in these decisions (respectively 131, 162, 82, 65, 49, and 50 in the same years). In effect, this shows that the *Autorité* has focused its enforcement on more serious offences and has, at the same time,

increased the fines it imposes, both on average and, in particular, in the case of the most serious offences.

2. Fines are Also Getting More proportionate to the Harm Caused by Cartels to the Economy

Do the above trends, that have been endorsed by the courts at the European level, and triggered by the legislature itself at the national level, mean that corporate fines have become excessive and/or that they will follow an ever-increasing pattern—as a few lawyers have suggested after throwing out a few figures? The issue can be addressed from at least two different angles: by looking at their aggregate level and by looking at their individual amount.

The claim that the aggregate amount of antitrust fines that has been levied in Europe (by the European Commission and by national competition authorities) in recent years is excessive generally rests on a comparison with figures in other jurisdictions, notably in the United States. Such a comparison is difficult to understand, not least because it sidesteps the fact that the main tool used to punish and deter cartelists in the United States is sending them to jail, a tradition that cannot be understood if one forgets that cartels have been legally considered as a felony since 1974 and are officially considered as white-collar economic crimes.³⁹ If one takes into consideration not only the significant and sustained increase of American corporate fines during the last half-decade (\$350 million in 2004, \$338 million in 2005, \$473 million in 2006, \$630 million in 2007, \$701 million in 2008, and \$1 billion in 2009), but also the significant and sustained increase of the number of days in jail to which guilty executives have been sentenced over the same period (7,334 days, 13,157 days, 5,383 days, 31,391 days, 14,331 days, and 25,396 days, respectively),⁴⁰ U.S. antitrust enforcement still appears to be considerably more severe than European antitrust enforcement. This conclusion is consistent with the finding that, on average, the penalty imposed in the United States on corporations alone is comparable to the fine imposed, on average, by the European Commission (respectively \$44 million and EUR 46 million for the period 2005/2009).

As for antitrust enforcement in other parts of the world, it is true that figures to date are undeniably lower. But then, which level is the “right” level? Are comparatively younger antitrust enforcement regimes relatively more lenient, or are comparatively older regimes overstretching themselves? History would tend to show that fines have only increased over time in the older regimes, and thus suggest that the younger regimes could follow a similar path in the coming years and decades. But, arguably, such an evolution would not tell us for sure who is right and who is wrong.

ON AVERAGE, THE PENALTY
IMPOSED IN THE UNITED STATES
ON CORPORATIONS ALONE IS
COMPARABLE TO THE FINE
IMPOSED, ON AVERAGE, BY THE
EUROPEAN COMMISSION.

This is why antitrust economists look, not at the aggregate level of corporate fines, but at the amount of the fine imposed in the case of each individual antitrust offence. Not being an economist myself, I will defer to what is currently the wisdom shared by eminent independent antitrust economists on this matter. The main studies in this regard⁴¹ agree that corporate fines better achieve deterrence than before, in that they are more proportionate to the illicit benefits that antitrust offenders are likely to expect when they enter into anticompetitive agreements as well as to the economic harm that such offences are likely to cause to consumers as a whole as well as to the broader economy.

This conclusion is consistent with the findings made by the *Autorité* in the course of the case-by-case assessment of the economic harm that it is legally bound to perform.⁴² However, most economists also concur on the fact that fines

THIS IS PRECISELY WHAT BOTH THE
LAW AND ECONOMICS OF
ANTITRUST WOULD VIEW AS A
COMBINATION OF
PROPORTIONALITY AND
DETERRENCE, I.E. MAKING FIRMS
UNDERSTAND THAT NOT
COMMITTING A SERIOUS
ANTITRUST OFFENCE IN THE FIRST
PLACE IS THE BEST POSSIBLE WAY
NOT TO END UP PAYING AN
EQUALLY SERIOUS FINE.

are still substantially lower than what would be needed to fully ensure deterrence if one takes into account not only the potential economic harm attached to cartels and other violations of antitrust law, as I have done up until now, but also their still limited rate of detection—in particular when they are covert in nature. I will come back to this issue when discussing criminalization below.⁴³

Like all economic assessments, these studies rest on assumptions that can be discussed to some extent. They nonetheless would seem to meet the “Daubert criteria” on the relevance and reliability of expert testimony,⁴⁴ meaning that they come from independent experts, that they have been published in scientific reviews after having been submitted to peer review, that they rest on transparent hypotheses as well as on scientifically accepted methods, and that they produce empirically tested and verifiable results. In any case, I am not aware of them having been challenged, to date, by other independent economists using equally or more robust tests and data.

To be sure, the statement recently made by a representative of the French antitrust bar in the mainstream economic press, to wit: “given the financial risk, it is better for a firm to breach its tax obligations, to commit an insider trading or to engage into money laundering than to fall in the hands of [antitrust enforcers],”⁴⁵ could be read as a confirmation that antitrust corporate fines are starting to become more than just “a cost of doing business” (in other words an economically sensible expense when compared to the huge profits that a cartel or sometimes an abuse of dominance can generate). This is precisely what both the law and economics of antitrust would view as a combination of proportionality and deterrence, i.e. making firms understand that not committing a serious

antitrust offence in the first place is the best possible way not to end up paying an equally serious fine.

That fines were not at adequate deterrence levels in the past, and even in the recent past, is difficult to doubt given the amazing rate of recidivism evidenced by antitrust enforcers. John Connor, in a recent paper that is perhaps the most comprehensive on the issue, studies a sample of international or regional cartels discovered in the last 20 years and finds a total of at least⁴⁶ 389 recidivists among firms found guilty of such an offence. He also finds that, although the mean number of cartels per recidivist is four, 52 corporations were members of seven or more cartels, 26 entered ten or more, and 6 engaged in twenty or more. Most strikingly, he stresses that, while in a number of occurrences firms that violated the law during the 1990's exhibited a slowing rate of recidivism in the 2000's (a period when more cartels were uncovered than during the previous decade), for most of the world's top antitrust recidivists the reverse occurred. It is precisely in this context, and bearing in mind the magnitude of the overcharge and broader negative welfare effects attached to cartels, that a number of competition agencies on both sides of the Atlantic increased corporate fines.

The conclusion that fines are now more proportionate to the seriousness and likely economic consequences of the offences that have been committed does not mean that these criteria are the only elements for competition authorities to take into account. The situation of each individual offender is clearly as important, especially at a time where corporations may be going through serious economic and financial difficulties which antitrust enforcers are not at all interested in making worse. But one of the main lessons of the last year is precisely that individual difficulties, when evidenced, are best handled in a tailor-made fashion, and not with a blanket curb put on corporate fines.

ONE OF THE MAIN LESSONS OF THE LAST YEAR IS THAT INDIVIDUAL DIFFICULTIES, WHEN EVIDENCED, ARE BEST HANDLED IN A TAILOR-MADE FASHION, AND NOT WITH A BLANKET CURB PUT ON CORPORATE FINES.

3. Fines are Systematically Individualized on a Case-by-Case Basis

Following the path of a number of other jurisdictions, starting with the United States and then the European Union, the *Autorité* has just announced its intention to publish the details of the guidelines it applies when setting corporate fines.

This guidance will not be the first to be released by a National Competition Authority (“NCA”) of the European Union. A number of NCAs—at least eight—have already done so during the last decade. Actually, the club to which they all belong, together with the European Commission, had itself agreed in 2008 on common principles intended to facilitate the convergence and consistency of fining practices throughout Europe,⁴⁷ which the *Autorité*'s draft duly takes into account.

But this draft will be the first guidance to be published in bad economic times. This context has led the *Autorité* to look very carefully at issues that had received relatively limited attention during the past decade of economic expansion, in particular the difficult issue of ability to pay. The ability of undertakings to pay the final fine must be thoroughly assessed. It is not the goal of competition authorities to make companies bankrupt because of having to pay an antitrust fine. At the same, the moral hazard linked to the fact that corporations that have broken the law, sometimes in a very severe manner, could have an interest in pleading an inability to pay without true justification in order to escape the fine, must not be overlooked either. The standard put forward in the draft published by the *Autorité* in March 2011 intends to balance both requirements. That being said, the *Autorité*'s fining guidelines are intended for good economic times as well

IT IS NOT THE GOAL OF
COMPETITION AUTHORITIES TO
MAKE COMPANIES BANKRUPT
BECAUSE OF HAVING TO PAY
AN ANTITRUST FINE.

as bad ones, meaning that the text must be flexible enough to accommodate all market situations as well as all individual situations.

The draft fining guidelines will also be the first in Europe, to my knowledge, to be released pursuant to a fully-fledged public consultation, which was launched on January 17, 2011 and lasted two months, until March 11. The draft that stakeholders had been invited to comment upon provided a comprehensive overview of the *Autorité*'s past and current fining practices, as well as of the case-law of review courts and of the European courts. It also revealed, for the first time, the different steps of the method followed in practice when assessing the various criteria provided by the law, and refined this method on a number of items. In doing so, it dwelled on European best practices, while at the same time incorporating the characteristics of French law which, as I state earlier, mandates a qualitative assessment of the harm caused to the economy. It also incorporated internationally accepted standards, including using a percentage point of the volume of affected commerce (or affected sales) as a base amount, depending on the seriousness of the infringement and on its likely economic impact, before taking into account individual elements relating to each offender's behavior and situation. Leaving aside leniency applicants who qualify for a full immunity, the draft then integrated rebates granted in case of a partial immunity or of a settlement. A special section was devoted to how each firm's ability to pay the fine at the time of the decision is assessed, as alluded to earlier on.

The criterion of the economic harm likely to flow from the offence warrants a few words, because it is rather specific to French antitrust law. The Code of Commerce and the case-law of the French Supreme Court do not require the *Autorité* to quantify the harm caused by a cartel or by an abuse of dominance, both for legal and policy reasons (public antitrust enforcement aims at punishing the offender and deterring it and other corporations from breaking the law, in the public interest, and not at measuring and compensating individual or collective

prejudices) and practical reasons (the offence has indeed taken place and it is impossible to objectively know for sure how things would have evolved precisely had it not occurred). What the legal provision on the determination of corporate fines requires from the *Autorité* is to prove its importance (in other words its order of magnitude), by making an assessment of the relevant qualitative and quantitative characteristics of the relevant market and broader economic context, on the basis of reasonably available data (aggregate market share of the offenders, barriers to entry, price-elasticity, actual or potential effects on competitors, and so on). The assessment of the actual or potential overcharge is an integral part of this exercise, but it is only a part of it, given that the law refers broadly to “the importance of the harm caused to the economy” and not strictly to “the extent of the overcharge,” as explained in greater detail at a recent session of the OECD’s competition committee.⁴⁸ This assessment, as I said earlier on, is a sometimes very demanding exercise, but it helps greatly in evidencing the proportionality of the final fine, among other elements to be taken into account, to its actual or potential economic incidence.

All along the process, the draft took care to achieve a balance between the three pillars that, I think, ground and should ground the sentencing practice of each and every competition authority: proportionality, individualization, and deterrence. These three aspects, it seems to me, cover the entire scope of issues that need to be addressed in the course of setting corporate fines, and are flexible enough to allow antitrust enforcers to take into consideration the specifics of each offence and of each individual offender, including the latter’s ability to pay the fine at the time it is due. I see no compelling reason for throwing the old-age wisdom that antitrust fines shall punish and deter, that they shall be individualized, and that they shall be proportionate to the offence at stake in the garbage of history, although I see a compelling reason for always updating this judicial legacy and adapting it to new events.

To make things square, I would add a fourth guideline to the three already cited: consistency, especially in the European Union, which is an integrated economy where all National Competition Authorities apply European antitrust law in addition to their domestic law in each case that is liable to substantially affect interstate commerce. In fact, consistency should not only be a policy goal; it is now a legal requirement.⁴⁹

The public consultation on the *Autorité*’s draft fining guidance has attracted wide attention, not only in France, and has resulted in almost thirty contributions that are expected to be published on the website of the *Autorité* at the same time as the final text, as is traditionally the case. It has given the agency the

ALL ALONG THE PROCESS, THE DRAFT TOOK CARE TO ACHIEVE A BALANCE BETWEEN THE THREE PILLARS THAT, I THINK, GROUND AND SHOULD GROUND THE SENTENCING PRACTICE OF EACH AND EVERY COMPETITION AUTHORITY: PROPORTIONALITY, INDIVIDUALIZATION, AND DETERRENCE.

opportunity of hearing a range of points of view and, in particular, a number of points of view that have few opportunities of making themselves heard. Besides business organizations and antitrust lawyers and economists, consumer associations, renowned academics, bar associations for other jurisdictions as well as other competition authorities have submitted very rich comments. These contributions, with many expressing very different positions, will tremendously help the *Autorité* in making its initial draft more precise, in incorporating new elements in the final guidelines, and in refining the overall balance of the document as well. We expect a final text to be published in the course of May 2011.

B. ARE CORPORATE FINES MISCONCEIVED IN ESSENCE AND SHOULD THEY BE REPLACED BY (OR COMBINED WITH) INDIVIDUAL PENALTIES?

The previous section leads me to share a number of conclusions reached either by Douglas Ginsburg and Joshua Wright, or by Joseph Harrington, or by both, in their recent discussion on antitrust sanctions.⁵⁰ But not all of them, though. The conclusions with which I agree are described in the following section.

1. Individual Compliance Matters as Well as Corporate Compliance With Antitrust Law

First, antitrust offences still abound and are probably still insufficiently deterred by corporate fines (although the recent mobilization of a number of corporations actually fined in the recent past in Europe and/or in the United States would tend to indicate that they now represent more than just “a cost of doing business”).

Second, corporate fines are probably unlikely to succeed in achieving deterrence alone (although they have been raised to a level which is now more proportionate to the harm that antitrust offences are liable to cause). This is especially so given the current context of economic recession.

Third, there are indeed a variety of persons involved in an antitrust offence: at least one legal person (the corporation(s)) and at least one physical person (the director(s) and/or employee(s)). I would stress, as do Douglas Ginsburg and Joshua Wright in their paper, that the physical persons act “within and on behalf” of the corporation, but I would immediately add that this works both ways. By that, I mean that there are two possible categories of persons who may have an incentive to violate antitrust law (or not) and who must be deterred from actually breaching it. My view is, therefore, that antitrust enforcement must “walk on two legs,”⁵¹ as I have said on many occasions, and that public enforcers should have in their toolbox instruments that allow them both to deter corporations from becoming antitrust offenders and to punish them when they actually become outlaws, and to deter and punish individuals. I would therefore complement corporate fines with individual penalties, and not substitute one with another.

A more radical shift could be advocated, that would in effect mean transferring the weight of deterrence from corporate fines to individual penalties alone. In support of this move, it is said notably that corporate fines end up hurting innocent corporations (and ultimately its innocent shareholders). I personally find it difficult to consider that firms can claim property of the profits made by their directors and employees in the course of business, including those generated by a prohibited business arrangement, but could at the same time disclaim liability for these same infringements in the event that they are discovered and prosecuted. Either you believe that corporations are only a set of individuals and indeed only those individuals can be held liable for their deeds in the same way as only they can benefit from their good actions. Or you accept that a legal person, in the same way as a physical person, has rights and liabilities, including liability linked to what the people who work for them do in the course of their business. Has any corporation whose director or employee was found guilty of an antitrust offense offered to hand this person back the profits generated by this offence before it was discovered and fined?

I therefore fully agree with the very insightful idea put forward by Douglas Ginsburg and Joshua Wright that we should have a somewhat more “granular” vision of what a corporation actually is and of how it actually operates, while adding that we should at the same time not lose sight of “conventional wisdom.” Corporate rights and duties exist, and it makes sense to align both corporate incentives (and shareholder incentives) and individual incentives (at directors’ level as well as at employees’ level) in complying with antitrust law rather than in breaching it or simply in not caring about it. This approach seems to be the one on which the European Parliament has settled on, as evidenced by its recent resolution on European competition policy in which it advocates “a wider range” of enforcement tools, including not only corporate “penalties that serve as an effective deterrent, in particular for repeat offenders,” but also “individual responsibility” and “compliance.”⁵²

Fourth, I would follow up on Joseph Harrington’ view on the “jail and/or debarment issue” by saying that the full mix of criminal penalties is useful and probably warranted in order to achieve individual deterrence. I would therefore add the faculty of debarring/disqualifying guilty individuals in appropriate cases to the possibility of sentencing them to jail terms in the case of a serious offence, rather again than substituting the latter with the former.

Fifth, I would add to the above the idea, also addressed by Douglas Ginsburg and Joshua Wright in their paper, that compliance can be driven forward by the traditional “stick and carrot” approach and that a serious compliance policy, that

I AGREE THAT WE SHOULD HAVE A SOMEWHAT MORE “GRANULAR” VISION OF WHAT A CORPORATION ACTUALLY IS AND OF HOW IT ACTUALLY OPERATES, WHILE ADDING THAT WE SHOULD AT THE SAME TIME NOT LOSE SIGHT OF “CONVENTIONAL WISDOM.”

is officially supported by the firm's philosophy and leadership, as well as actually implemented and monitored and accompanied by sanctions in case of breach, can go a long way in bringing on board individual incentives for antitrust compliance. This is, in fact, a topic on which the *Autorité de la concurrence* has committed itself, by announcing that it would release a draft policy document on antitrust compliance in autumn 2011 and, at the same time, launch a public consultation on draft guidelines on antitrust settlement.⁵³

Sixth, the experience of the *Autorité* is that the toolbox of antitrust enforcers can usefully comprise injunctions of publication, whereby the agency requires guilty firms to publish summaries of the case in the general or special media. This power is provided for by French law⁵⁴ and is routinely used by the *Autorité*, notably in cartel cases. It has obvious enforcement and advocacy virtues.

2. Making Criminal Enforcement Effective in the French Context

Creating a criminal antitrust law (where it does not exist at present) and/or beefing up criminal antitrust enforcement (where a law already exists as is the case in France) is one issue. Articulating it smoothly with administrative enforcement is another. I will turn to each of these issues briefly.

I AM A SUPPORTER OF CRIMINAL
ANTITRUST ENFORCEMENT AGAINST
INDIVIDUALS, BUT I ALSO
THINK THAT THIS TYPE OF
ENFORCEMENT SHOULD REMAIN
WHERE IT BELONGS.

I am a supporter of criminal antitrust enforcement against individuals, but I also think that this type of enforcement should remain where it belongs: in the hands of criminal judges and public prosecutors, while independent competition authorities remained firmly focused on

enforcement against corporations. In my mind, given the constitutional, legal, and historical background of a number of Member States of the European Union, and in particular of France, these are really seen as being two very different, albeit complementary, jobs.

Obviously, however, just having a criminal law in place does not mean that it will be applied, as is the case in France. This situation is unfortunate and I want to shed some light on why this is so and what can be done to make progress.

First, there is certainly a problem in the law itself. In most countries where it exists, such as the United Kingdom or the United States, antitrust criminal law is focused on secret, hardcore offences (cartels and bid-rigging). This, it seems to me, makes sense from an economic viewpoint (limiting risks of wrong conviction), from a legal standpoint (ensuring certainty and predictability), and as a matter of policy (focusing public morality). Only hardcore offences, defined as those that unmistakably hurt the economy and consumer welfare and, by and large, lack any redeeming virtue—characteristics that make their legal status

clear-cut and which everyone consequently understands upfront are wrong—fall in that category.

Other antitrust offences are not so clear-cut and cannot consequently be understood upfront to be wrong. This is because such offences, including most vertical restraints and abuses of dominance other than practices blatantly aimed at excluding or exploiting competitors, are not unmistakably liable to harm competition and incapable of producing economic efficiencies.

In France, however antitrust law makes it a criminal offence for individuals to participate, in a decisive and fraudulent way, in any type of anticompetitive practice, either in combination or alone. The fact that the current provision is so broad and thus not so clear-cut is certainly one of the reasons why it is rarely applied to date. This leads me to advocate not “more criminalization,” but “better criminalization,” by narrowing down the scope of criminal law to hardcore anticompetitive practices.

Second, there are a number of problems with the enforcement of the law. To start with, public policy-makers currently do not prioritize “crime in the suit,” as it has been described in the United States⁵⁵ as much as “crime on the street.”

But this de-prioritization could be rethought given the diagnosis that the 2008/2009 financial crisis, and the ongoing economic recession that it has engendered, was caused by excessively risky and sometimes unconscionable business practices that were left unchecked by public policy-makers until it was very late in the day. If we roll up our sleeves to re-dimension the financial regulatory framework, and to re-mobilize public bodies in charge of applying it, should we not do the same in the field of antitrust? If we agree that we need not only adequate rules and dedicated agencies in charge of making sure that companies play by these rules, but also public prosecutors committed to making sure that individuals comply with the law as well, should we not apply this lesson to antitrust as well?

In addition, although judges have become specialized to some extent, they still lack the expertise and experience needed to make them familiar with antitrust criminal enforcement. Training would help, in order to make sure that the cases brought by public prosecutors are well-understood and well-handled.

Third, there are certainly ways of better coordinating prosecutors and judges in charge of criminal enforcement against individual offenders and competition authorities in charge of administrative enforcement against corporate offenders. I’ve already underlined the most important of them all in my mind: putting in place a program of individual leniency, in parallel to the one that already exists for firms.⁵⁶ This would not only benefit directors and employees who decide to help competition authorities to detect and successfully prosecute cartels by giv-

THIS LEADS ME TO ADVOCATE NOT “MORE CRIMINALIZATION,” BUT “BETTER CRIMINALIZATION,” BY NARROWING DOWN THE SCOPE OF CRIMINAL LAW TO HARDCORE ANTICOMPETITIVE PRACTICES.

ing them the benefit of a full or partial immunity against debarment or jail. It would also benefit companies. Today, companies which contemplate cooperating with a competition enforcer in exchange for a fine immunity or reduction often struggle to secure the help of their current or past directors and employees, who are left unprotected against the risk of a jail term. Tomorrow, with an individual leniency program in place, the incentive of directors and employees to cooperate would be better aligned with the incentive of their firms to do the same.

This small proposal, which could have a big outcome while not costing a cent, is not revolutionary. It has specifically been advocated in a recent report on the future of French business law,⁵⁷ and could be readily incorporated in a draft bill on the modernization of economic justice.

IV. Going Forward

To sum up the above, I would remind ourselves that in Europe, as well as in most parts of the world, the number one priority of citizens is to get their fair share of the market economy that they are asked to support. In practice, this means (1) getting a job and (2) earning sufficient money to be able to rest, to support their family, and to plan for their future. This is why competition and competition law and policy are directly relevant to their wellbeing. They are the Magna Carta of the market economy because they drive firms and, when needed, remind them not only to do their best on the marketplace but also to do it for the benefit of consumers (who are their customers as well as, often, their employees), rather than at the expense of consumers.

In a time of economic globalization and recession, as well as of rising prices (especially for commodities, energy and food), competition is a good deal for corporations and consumers alike. Promoting competition law and enforcement, and making it more efficient rather than relaxing it or abandoning it, should therefore also be a good deal for policy-makers. ▼

-
- 1 Bruno Lasserre, *How Can National Competition Authorities Mobilize in Times of Global Crisis?*, 12(1) GLOBAL COMPETITION POL'Y., December 15, 2008 (available online at: <https://www.competitionpolicyinternational.com/how-can-national-competition-authorities-mobilize-in-times-of-global-crisis/>).
 - 2 Decision n° 09-DCC-16 of the Autorité de la concurrence of 22 June 2009 relating to the merger between Caisse d'Épargne and Banque Populaire (available online at: http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=316&id_article=1234)
 - 3 This debate seems to surface each time a political or economic crisis appears. Enlightening, although very different, perspectives on its past occurrences can be found in Carl Shapiro, *Competition Policy in Distressed Industries*, Remarks as prepared for delivery to the ABA Antitrust Symposium on Competition as a Public Policy, May 13, 2009 (available online at: <http://www.usdoj.gov/atr/public/speeches/245857.htm>) and Daniel A. Crane, *Antitrust Enforcement During National Crises: An Unhappy History*, 12(1) GLOBAL COMPETITION POL'Y., (Autumn 2008), available online at: <https://www.competitionpolicyinternational.com/antitrust-enforcement-during-national-crises-an-unhappy-history/>).

- 4 The projects undertaken by the ICN in this respect, as well as the products delivered notably within the Merger Working Group and the Unilateral Conduct Working Group, are all *available* online at: (<http://www.internationalcompetitionnetwork.org/library.aspx>).
- 5 For a mid-term assessment of what has been done and what remains to be done, see Bruno Lasserre, *The New French Competition Law Enforcement Regime*, COMPETITION L. INT'L, (October 2009) (*available* online at: http://www.autoritedelaconcurrence.fr/doc/competition_law_international_october_2009.pdf).
- 6 Specifically, Article 101 of the Treaty on the Functioning of the European Union (TFEU) states that agreements prohibited because of their anticompetitive object or effects can nonetheless be justified where they "contribute to improving the production or distribution of goods or to promoting technical or economic progress" (efficiency), "while allowing consumers a fair share of the resulting benefit" (fairness). In essence, a similar justification is available in the case of *prima facie* abusive unilateral conduct prohibited by Article 102 of the TFEU. The test and formula are similar under French law (see Article L. 420-1, L. 420-2 and L. 420-4 of the Code of Commerce).
- 7 Barak Obama, *Statement for the American Antitrust Institute*, 2007 (*available* online at: http://www.antitrustinstitute.org/files/aaai-%20Presidential%20campaign%20-%20Obama%209-07_092720071759.pdf).
- 8 The legislative package, which comprises two main pieces (Law No. 2008-776 of 4 August 2008 on the Modernization of the Economy and Ordinance No. 2008-1161 of 13 November 2008 on the Modernization of Competition Enforcement), includes a number of other measures intended to make the economy more competitive, initially suggested by the Commission on the Liberation of French Growth in its report of January 2008 (*available* online at: <http://lesrapports.ladocumentationfrancaise.fr/BRP/084000041/0000.pdf>).
- 9 See references, *supra* note 3.
- 10 Both in domestic policies and on the international scene (see in this regard the letter from the President of the United States of September 6, 1944 to the Secretary of State concerning cartel policy).
- 11 Mario Monti, *A New Strategy for the Single Market—At the Service of Europe's economy and society*, Report to the President of the European Commission, 9 May 2010 (*available* online at: http://ec.europa.eu/internal_market/strategy/docs/monti_report_final_10_05_2010_en.pdf).
- 12 *Appalachian Coals v. US*, 288 U.S. 344 (1933).
- 13 Judgment of the Court of Appeal of Paris of 19 January 2010, *AMD Sud-Ouest a.o.*
- 14 The Minister for Economy, who alone had the power to challenge the case before the Cour de Cassation (Supreme Court), did not do so at the time. The Parliament regretted this situation and empowered the Autorité de la Concurrence to make such an appeal in future cases. In a recent judgment, the Supreme Court clearly distanced itself from the Court of Appeal.
- 15 Recommendation No. C(98)35/Final of the Council of the OECD of 25 March 1990 concerning Effective Action against Hard Core Cartels.
- 16 Judgment of the European General Court of 15 July 1994 in Case T-17/93, *Matra Hachette/European Commission*.
- 17 Judgment of the European Court of Justice of 16 December 1975 in Joined Cases 40/73 to 48/73, 50/73, 54/73 to 56/73, 111/73, 113/73 and 114/73, *Suiker Unie a.o./European Commission*.

- 18 Judgments of the European Court of Justice of 13 July 1966 in Joined Cases 56/64 and 58/64, Consten & Grundig, and of 6 October 2009 in Joined Cases C-501/06 P e.a., GlaxoSmithKline/European Commission.
- 19 Judgment GlaxoSmithKline/European Commission, *Id.* ¶163: “[i]t must be borne in mind that the Court has held that, like other competition rules laid down in the Treaty, Article [101] aims to protect not only the interests of competitors or of consumers, but also the structure of the market and, in so doing, competition as such. Consequently, for a finding that an agreement has an anti-competitive object, it is not necessary that final consumers be deprived of the advantages of effective competition in terms of supply or price.” A comparable philosophy grounds U.S. antitrust law as interpreted in a number of past cases by the U.S. Supreme Court.
- 20 With very negative consequences for the economy as a whole, as evidenced by studies on the tolling of antitrust in the United States during the Great Depression of the 1930’s. See for instance Howard A. Shelanski, *Enforcing Competition During an Economic Crisis*, in Symposium: The Effect of Economic Crises on Antitrust Policy, 77(1) ANTITRUST L.J. (2010).
- 21 See my recent hearing before the Parliament (record of session No. 78 of 23 June 2010 of the Economic Committee of the French National Assembly, *available* online at: http://www.assemblee-nationale.fr/13/cr-eco/09-10/c0910078.asp#P6_217), as well as “*Europe must defend its interests while remaining faithful to its openness to the world*”, joint tribune published in LE MONDE of 9 February 2011 by Mikolaj Dowgielewicz, Werner Hoyer, Diego Lopez Garrido, Pedro Lourtie, Paolo Romani, & Laurent Wauquiez, European Secretaries of State or Ministers of Economy of Poland, Germany, Spain, Portugal, Italy, and France.
- 22 DAVID GERBER, GLOBAL COMPETITION: LAW, MARKETS & GLOBALIZATION, (2010).
- 23 Agreement on Subsidies and Countervailing Measures (*available* online at: http://www.wto.org/english/docs_e/legal_e/24-scm.pdf).
- 24 Article 107 of the Treaty on the Functioning of the European Union.
- 25 For a comprehensive and insightful analysis of the criminalization issue in the context of European antitrust law and policy, see Wouter P.J. Wils, *Is Criminalization of EU Competition Law the Answer?*, EUROPEAN COMPETITION LAW ANNUAL 2006: ENFORCEMENT OF PROHIBITION OF CARTELS (Claus-Dieter Ehlermann & Isabela Atanasiu eds.) (2006).
- 26 Article L. 420-6 of the Code of Commerce.
- 27 Judgment of the European Court of Justice of 7 June 1983 in Joined Cases 100 to 103/80, Musique Diffusion Française/European Commission.
- 28 Guidelines of the European Commission of 1st September 2006 on the method of setting fines imposed pursuant to Article 23(2)(a) of Regulation No. 1/2003 (*available* online at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:210:0002:0005:EN:PDF>).
- 29 Guidelines of the European Commission of 14 January 1998 on the method of setting fines imposed pursuant to Article 15(2) of Regulation No. 17 (*available* online at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:1998:009:0003:0005:EN:PDF>).
- 30 European Commission press release No. IP/06/857 of 28 June 2006, *Commission revises guidelines for setting antitrust fines* (*available* online at: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/06/857&format=HTML&aged=0&language=EN&guiLanguage=en>).

- 31 Prior to judicial review (available online at: <http://ec.europa.eu/competition/cartels/statistics/statistics.pdf>).
- 32 Law N. 2001-420 of 15 May 2001 on New Economic Regulations.
- 33 Article L. 464-2 of the Code of Commerce.
- 34 I have looked in greater detail at this issue in a previous article: *La regulation concurrentielle un an après sa réforme: un point de vue d'autorité (part 2)*, (4) CONCURRENCES (2010) (available online at: http://www.autoritedelaconurrence.fr/doc/partie2_unanap_reforme_bl_concurrences.pdf).
- 35 Prior to judicial review.
- 36 Due mainly to a case in which the *Conseil de la concurrence* fined a nationwide cartel committed by the three domestic mobile phone operators (decision No. 05-D-53 of 30 November 2005 relating to practices committed in the mobile telephony sector, available online at: http://www.autoritedelaconurrence.fr/user/standard.php?id_rub=160&id_article=502).
- 37 Due mainly to a case in which the *Conseil* fined a nationwide cartel committed by steel trade operators (decision No. 08-D-32 of 16 December 2008 relating to practices committed in the steel trade industry, available online at: http://www.autoritedelaconurrence.fr/user/standard.php?id_rub=256&id_article=1015), an offence which the Paris Court of Appeal confirmed on the merits, while very sharply reducing the fines (*supra* note 12). The *Conseil* did not legally have the power, at the time, to appeal judgments of the Paris Court of Appeal before the Supreme Court. This power has been provided for since then.
- 38 Due mainly to a case in which the *Autorité* fined 11 banks for colluding on fees charged on the processing of checks (decision No. 10-D-38 of 20 September 2010 relating to the tariffs and condition on the handling of checks, available online at: http://www.autoritedelaconurrence.fr/user/standard.php?id_rub=368&id_article=1472).
- 39 Scott D. Hammond, *The Evolution of Criminal Antitrust Enforcement Over the Last Two Decades*, 24th Annual National Institute on White Collar Crime, February 25, 2010.
- 40 *Id.*
- 41 Notably the numerous studies published by John Connor on American and European antitrust enforcement, as well as by Emmanuel Combe on European antitrust enforcement, as well as those cited therein.
- 42 See § II.A.2 above.
- 43 See § III.B below.
- 44 *Daubert v. Merrell Dow Pharms. Inc.*, 509 U.S. 579 (1993).
- 45 *Atteintes à la concurrence: pour des sanctions plus cohérentes*, LES ECHOS, 20 December 2010.
- 46 John M. Connor, *Recidivism Revealed: Private International Cartels 1990-2009*, 6(2) COMPETITION POL'Y INT'L, (Autumn 2010). The actual figure might be greater because some competition authorities do not reveal the name of fined offenders, or at least of some of them, in their published decisions.
- 47 European Competition Authorities (ECA), *Principles for convergence on pecuniary sanctions imposed on undertakings for infringements of antitrust law*, (May 2008) (available online at: http://www.autoritedelaconurrence.fr/doc/eca_principles_uk.pdf).

- 48 See note No. DAF/COMP/WD(2011)16 of OECD of 4 February 2001, *Roundtable on the quantification of harm to competition by national courts and competition agencies: note of France* (available online at: http://www.autoritedelaconurrence.fr/doc/ocde_quantif_harm_fev11_uk.pdf).
- 49 Judgment of the European Court of Justice of 11 June 2009 in Case C-429/07, *X BV*.
- 50 Douglas H. Ginsburg & Joshua D. Wright, *Antitrust Sanctions*, and Joseph Harrington, *Comment on Antitrust Sanctions*, 6(2) *COMPETITION POL'Y INT'L*, (Autumn 2010).
- 51 Notably during the ICN's 2009 Annual Conference (Zurich, Switzerland)'s Cartel Working Group Plenary.
- 52 Resolution No. 2010/2137 of 20 January 2001 of the European Parliament on the report on competition policy for 2009, ¶¶ 56, 60 (available online at: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2011-0023+0+DOC+XML+V0//EN>).
- 53 See press release of the *Autorité* of 17 January 2011.
- 54 Article L. 464-2 of the Code of Commerce.
- 55 *Supra* note 39.
- 56 Article L. 464-2 of the Code of Commerce.
- 57 Report of January 2008 of the Working Group presided by Jean-Marie Coulon on the Future of Business Law (available online at: <http://lesrapports.ladocumentationfrancaise.fr/BRP/084000090/0000.pdf>).



Using a Sledgehammer to Crack a Nut: Why China's Anti-Monopoly Law was Inappropriate for Renren v. Baidu

Angela Huyue Zhang

Cleary Gottlieb

Using a Sledgehammer to Crack a Nut: Why China's Anti-Monopoly Law was Inappropriate for *Renren v. Baidu*

By Angela Huyue Zhang*

On December 18, 2009, Beijing No. 1 Intermediate People's Court issued a ruling in favor of Baidu, Inc., a leading search engine provider in China, in an abuse of a dominant position case brought by Tangshan Renren Information Services Co., an operator of a medical information consulting website. Renren alleged that Baidu had downgraded its website in order to coerce it into using its search advertising services. The Court dismissed the case primarily on the grounds that Renren had failed to establish that Baidu had a dominant position in China's search engine service market.

Although the dismissal may have been the correct outcome, the Court's analysis was misguided. While the Court recognized certain two-sided features of Baidu's business model, it failed to further explore the impact of those features on the competition analysis. Crucially, the Court erred in defining the relevant product market as the search engine service market. Instead of using a one-sided approach, the Court should have adopted a two-sided approach in defining the relevant market.

Moreover, the Court readily accepted Baidu's defense without investigating whether the blockage was *solely* motivated by the existence of junk links. Indeed,

*Angela Huyue Zhang is an associate at Cleary Gottlieb Steen & Hamilton LLP in Brussels. The author would especially like to thank Judge Richard A. Posner, David S. Evans, and Gregory Sidak for invaluable comments and inspiration. She is also indebted to Alan Devlin, Su Sun, Song Yang, Lindsay McSweeney and Wentong Zheng for help and suggestions. The author takes full responsibility of the content and welcomes comments and criticisms. She can be reached at: angela.angelazhang@gmail.com

the information asymmetry between Baidu and customers such as Renren made it difficult to discern whether Baidu had downgraded the websites with the legitimate reason of penalizing junk links or with the motive of coercing those websites into using its advertising services.

On the other hand, although there is a theoretical possibility that Baidu may have had an incentive to impose artificial switching costs in order to lock in existing customers, the reputational cost should have been sufficient to deter Baidu from committing such abuses. New customers who are informed about the switching cost would be unlikely to choose Baidu and existing customers who are locked in would be unlikely to choose Baidu again. As informed consumers would not be harmed, the application of the Anti-Monopoly Law to this case is like using a sledgehammer to crack a nut. Indeed, consumer protection law rather than antitrust law would have been a better tool to tackle abusive behaviors like those alleged by Renren.

I. Introduction

Since China's Anti-Monopoly Law ("AML") took effect in August 2008, the Ministry of Commerce ("MOFCOM"), the agency in charge of merger control in China, has made most of the headlines relating to the AML through its handling of several high-profile merger cases.¹ By contrast, the State Administration for Industry and Commerce ("SAIC") and the National Development and Reform Commission ("NDRC"), the competition authorities responsible for enforcement against agreements among competitors and the abuse of dominant positions, have been slow to prosecute cases.² As a result, an increasing number of complainants are resorting to private lawsuits in the Chinese courts for remedies. Under Article 50 of the AML, operators who implement monopolistic conduct and cause loss to others shall bear civil liability according to law. This provision thus provides a statutory basis for private parties to bring suits under the AML.

So far a few cases have been reported, and one of the most high-profile cases involves Baidu, Inc. ("Baidu"), a leading Chinese language internet search provider. Baidu provides a paid for performance service (hereinafter "paid search advertising service"), a web-based auction system that allows advertisers to bid for positions in the relevant keyword search.³ Tangshan Renren Information Services Co. ("Renren"), an operator of a medical information consulting website, started using Baidu's search advertising service to promote its website ("qmyyw.com") in March 2008. Three months later, when Renren began to reduce its spending on the search advertising service, it immediately noticed a significant decrease of visits to its website. Renren then alleged that Baidu had blocked internet users' access to Renren's website in order to coerce Renren to use more of Baidu's search advertising services. Beijing No. 1 Intermediate People's Court (the "Court") accepted the case on January 6, 2009 and it went to trial on April 22, 2009. On December 18, 2009, the Court issued a ruling in favor of Baidu on the grounds that Renren had failed to meet its burden of proof in establishing that Baidu held a dominant position in China's search engine service market.⁴

AS THE FIRST DETAILED DECISION
ISSUED BY A CHINESE COURT,
RENREN V. BAIDU PROVIDES
IMPORTANT INSIGHTS INTO HOW
FUTURE PRIVATE LITIGATIONS
MAY PROCEED IN CHINA.

This case has received a tremendous amount of attention, both by economists and lawyers.⁵ As the first detailed decision issued by a Chinese court, *Renren v. Baidu* provides important insights into how future private litigations may proceed in China. Commentators applauded the decision as the Court required the plaintiffs to satisfy a high evidentiary threshold, which was regarded as consistent with international standards.⁶ While the importance of the case has rightly been recognized, some of the most fascinating and thorny economic issues have not yet been researched or analyzed. This article provides an in-depth analysis of the case by focusing on the following questions: First, what were the unique features of

Baidu's search advertising services? Second, how did those features affect competition analysis in this case? In particular, why did the Court err in defining the relevant product market as the search engine service market? Third, what were the problems in the Court's assessment of the alleged abusive behavior? And finally, would consumer protection law rather than antitrust law have been a better tool to tackle abuses like those alleged by Renren in this case?

II. The Search Advertising Services

Like Google, Yahoo! and other internet search engine providers, Baidu realizes its profits through search advertising services, which create a platform for advertisers to bid for priority placement of their links in keyword searches. Prior to September 2006, advertisers' links were ranked solely based on the price bid of that keyword.⁷ Since then, links to advertisers' websites have been ranked according to a comprehensive ranking index based on both the quality factor of a keyword and the price bid of that keyword.⁸ The quality factor of a keyword is mainly based on the relevance of the keyword, which is determined by an analysis of past searches and click-through results.⁹ Each time a searcher clicks on an advertiser's link in the search results, Baidu recognizes revenue based on the amount of the bidding fee that the advertiser has agreed to pay and the quality factor of the keyword.¹⁰

While Baidu claims to be the first auction-based search advertising service provider in China,¹¹ Overture Services (formerly GoTo.com, now part of Yahoo!) was the first to introduce the auction system to sell sponsored search advertising and, since its introduction in 1998, this system has been used widely by worldwide search engine providers around the world.¹² One well-known example is Google, which uses a similar auction-based system called "Google Adwords" to sell advertising space. However, at the time of this case, there was a stark difference in the display of paid search results between Baidu and Google. Google's search advertising clearly segregates natural search results (also called "organic search results") from paid search results: The natural results appear on the left-hand side and the paid search results mainly appear on the right-hand side. Although the paid search results sometimes appear on top of the natural search results on the left-hand side, they are in shadow and thus are clearly separated from the natural search results.

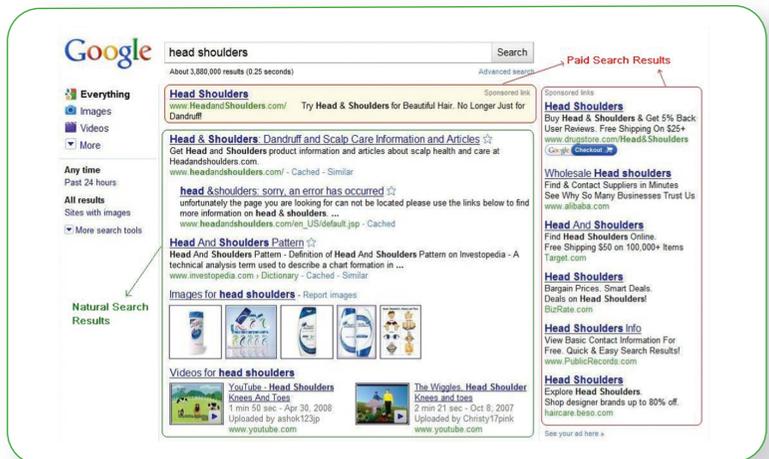
Notably, creating a clear distinction between paid search results and natural search results was recommended by the United States Federal Trade Commission (the "FTC"). In 2002, in response to a complaint that search engine providers were violating Section 5 of the FTC Act by failing to disclose that advertisements were inserted into search engine results lists, the FTC sent out a letter to all search engine providers recommending that their websites use clear

NOTABLY, CREATING A CLEAR
DISTINCTION BETWEEN PAID
SEARCH RESULTS AND NATURAL
SEARCH RESULTS WAS
RECOMMENDED BY THE
UNITED STATES FEDERAL
TRADE COMMISSION.

and conspicuous disclosures to ensure that any paid ranking search results be distinguished from natural search results.¹³

In contrast to Google's and other U.S. search engine providers' practices, Baidu mixed its natural search and paid search results. Before the development of Phoenix Nest,¹⁴ a keyword auction system introduced by Baidu in November 2008, the natural search results were commingled with the paid search results, both appearing in the same search results list. The only distinction between the two was that natural search results were marked with "Baidu Quick Webshot," while paid search results were marked as "Promotion." (The comparison of Baidu and Google's display of search results is illustrated in Figure 1 below.)

Figure 1



By mixing natural search results with paid search results, Baidu, in effect, offered advertisers a means to manipulate search results through the search advertising services. Such a practice undermined the reliability of Baidu's search

engine service. Although the paid search results were marked differently from natural search results, less sophisticated internet users may not have readily discerned the significance of the meanings of “*Baidu Quick Webshot*” and “*Promotion*.” Some internet users may not even have noticed such a subtle distinction and may have been misled into believing that the paid search results were natural search results. Indeed, Baidu’s strategy of mixing paid search results with natural search results significantly contributed to Baidu’s rapid growth: The promotional links on the left-hand side generated many more click-throughs and thus brought in more revenue than promotional links on the right-hand side.¹⁵

The screen shot of Baidu’s website above illustrates this problem. Kaixin001.com is a highly successful social networking website (similar to Facebook) in China. A rival launched a competing service under a similar name of Kaixin.com and tried to clone Kaixin001. To attract more attention from internet users, Kaixin used Baidu’s paid search advertising services to boost its ranking on the search result list. As a result, less informed internet users may have been misled into believing that Kaixin was a popular website because it had a high ranking on the search result list and some may even have confused Kaixin with Kaixin001.

BY MIXING NATURAL SEARCH RESULTS WITH PAID SEARCH RESULTS, BAIDU, IN EFFECT, OFFERED ADVERTISERS A MEANS TO MANIPULATE SEARCH RESULTS THROUGH THE SEARCH ADVERTISING SERVICES.

Furthermore, since the ranking of the search results was not solely determined by relevance, even if internet users were informed about the distinction between paid search results and natural search results, this manipulation of search results increased the transaction costs incurred by internet users looking for relevant information. Unsurprisingly, although Baidu’s strategy was very successful, it was also very controversial in China.¹⁶

In the United States, there has been an intense academic debate on how to regulate search engines.¹⁷ As noted earlier, the FTC has recommended that search engines segregate paid search results from natural search results. However, Chinese regulators such as the Ministry of Industry and Information Technology (responsible for regulating the internet industry) and SAIC (responsible for consumer protection) have yet to initiate similar actions to limit the manipulation of search results.

III. The Court’s Opinion

Due to the importance of this case, the presiding judge, Tong Shu, read out the decision in a live broadcast. The transcripts of the live broadcast provide important insights into the Court’s application of the AML.¹⁸ Judge Shu later published an article on this case in an English journal in China, and her analysis in the article is substantially similar to that in the opinion.¹⁹

A. RENREN'S ARGUMENTS

Renren first argued that Baidu had gained a dominant position under the AML. It relied on two pieces of evidence: the first was an article from *China Securities Journal* reporting that a Beijing consulting firm had estimated that Baidu had a 65.8 percent market share in China's search engine market; the second was a news report on Baidu's own website claiming that Baidu had a market share of over 70 percent in China's search engine market. Applying Article 19 of the AML, which provides that an entity with a market share of 50 percent or more is presumed dominant, Renren argued that Baidu held a dominant position.

Renren then submitted a comparison of the results generated in searches for its website on both Google and Baidu's websites in September 2008. Upon entering a query for Renren's website, internet users were presented with 6,690 results on Google and 4 results²⁰ on Baidu. Renren thus accused Baidu of abusing its dominant position in violation of Article 17 (4) of the AML, which prohibits any "undertaking that requires a trading party to trade exclusively with itself or trade exclusively with a designated business operator without justifiable cause." Renren requested an award of RMB 1,106,000 (approximately \$162,000) for damages as well as the removal of the blockage of its website.

B. BAIDU'S RESPONSE

Baidu did not contest that it had blocked Renren's website. It argued, however, that it did so because Renren's website included a large amount of junk links, and that the search engine automatically penalized such a practice. Baidu alleged that the junk links were irrelevant to Renren's webpage and that Renren manually set up those links in order to boost its website's natural ranking. Baidu clarified that the blockage only applied to natural search results, and was not relevant to the amount of payment for the search advertising services.

Further, Baidu argued that Renren lacked substantive evidence in alleging that Baidu possessed market power. A search engine service provided free to internet users, Baidu claimed, could not be a "relevant market" under the AML. Besides, the argument continued, the evidence that Renren provided about Baidu's market share only related to a temporary, very short period; and, moreover, market share could not be used as the single standard in evaluating whether Baidu possessed market power.

In arguing that a free engine service could not be a relevant market, Baidu mainly relied on *KinderStart.com, LLC v. Google, Inc.* ("KinderStart"),²¹ an unpublished decision by the district court in the Northern District of California.²² On March 13, 2006, KinderStart filed a class action lawsuit against Google, alleging that the search engine company had illegally blocked a multitude of websites, including that of KinderStart. KinderStart sought relief based on a wide range of legal theories under state and federal law, including, among others, attempted monopolization and monopolization in violation of the Sherman Act.²³

KinderStart identified two relevant markets for the attempted monopolization and monopolization claims: the search market and the search advertising market.²⁴ In dismissing KinderStart’s claims, the court held that KinderStart had failed to establish the search market and the search advertising market as relevant markets.²⁵ In particular, the court noted that KinderStart had “cited no authority indicating that antitrust law concerns itself with competition in the provision of free services,” and therefore concluded, “the search market is not a ‘market’ for purposes of antitrust law.”²⁶

C. THE COURT’S REASONING

The first issue confronting the Court was the definition of the relevant market. The Court relied on Article 12(2) of the AML, which provides that “relevant market” refers to “the commodity scope or territorial scope within which the business operators compete against each other during a certain period of time for specific commodities or services.” The Court then cited Article 3 of the Guidelines for the Definition of Relevant Market promulgated by the Anti-Monopoly Commission, which provides the definitions, among others, of relevant product market and relevant geographic market. After examining the substitutability of the search engine service with other types of internet services (such as internet news services, instant messenger services, and other internet services), the Court held that the search engine service constitutes an independent relevant market. In addition, considering cultural differences, language preferences, and other factors, the Court decided it was appropriate to define the relevant geographic market as China. The Court then concluded that the relevant market was “China’s search engine service market.”

THE COURT HELD THAT THE SEARCH ENGINE SERVICE CONSTITUTES AN INDEPENDENT RELEVANT MARKET.

Meanwhile, the Court rejected Baidu’s argument that the search engine service itself could not be a relevant market because the AML does not apply to free services. The Court was unpersuaded by *KinderStart* and reasoned that although the search engine service was free, the service was closely tied to other products and services for which Baidu did require payment.²⁷ Unlike free public internet services, search engine services generate actual or potential profits from advertising and marketing. Therefore, whether a service is free is an irrelevant factor in evaluating the relevant market.

After defining the relevant market, the next issue the Court considered was the question of whether Baidu held a dominant position. The Court mainly relied on two provisions under the AML: Article 18, which provides a list of factors for evaluating the existence of a dominant market position; and Article 19, which establishes a rebuttable presumption of dominance in cases where a firm’s market share exceeds 50 percent. After examining the two news articles submitted by Renren, the Court held that such evidence was insufficient to establish Baidu’s market power. First, it asserted, the two news articles did not clearly

define the market as it was not clear whether the search engine market referred to in the articles exactly matched the relevant market defined by the Court. Second, both articles failed to provide sufficient information regarding the underlying data and method used to calculate Baidu's market share. Therefore, the Court was not convinced that the articles were based on scientific and objective analysis. After rejecting this evidence, the Court did not proceed further to investigate Baidu's market power.

In addition, the Court found that Baidu had a pro-competitive justification for downgrading Renren's website as Renren had attempted to increase the natural ranking of its website by adding many unrelated junk links. As Baidu's policy of prohibiting junk links had been published on its website, Renren could have been aware of the policy. Moreover, the policy applied to all websites and did not specifically discriminate against Renren's website. Importantly, the Court reasoned that Baidu's policy had legitimate value because it increased the accuracy and reliability of search results to the benefit of internet users. Moreover, there was no evidence that Baidu's practice was discriminatory or coercive to Renren; in fact, Baidu had sufficient legitimate reasons to block Renren's website. For these reasons, the Court denied Renren's request for damages and the removal of blockage.

IV. Analysis of the Case

A. DEFINING THE RELEVANT MARKET

In defining the relevant market, the Court employed a one-sided approach and identified the relevant product market as the search engine service market. This was a crucial mistake. Baidu is a two-sided platform that simultaneously serves both advertisers and internet users. Instead of using a one-sided approach, the Court should have adopted a two-sided approach to identifying the relevant market in this case. In fact, even viewed from a one-sided perspective, the Court erred in defining the relevant market. Renren's main allegation was that Baidu had coerced it into using its search advertising services; therefore, the Court should have focused on the advertising side rather than the search engine side to evaluate Baidu's market power.

Interestingly, in rebutting Baidu's argument that antitrust law did not apply to the search engine service market (because its services are offered free of charge), the Court seemed to recognize certain two-sided features of Baidu's business model—the search engine, it noted, is closely tied to other products or services for which Baidu does requires payment. Regrettably, the Court did not explore further how the two-sided features have a significant impact on competition analysis in this case.

Studies on two-sided markets date back to 1983 when William Baxter developed certain foundational insights on two-sided platforms in his treatment of the

payment card system.²⁸ In 2003, Jean-Charles Rochet and Jean Triole provided the first formal analysis in their pioneering work on two-sided markets:

“A market 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.”²⁹

Rochet and Triole use the term “market” in a loose manner that does not agree with how it is used in competition policy.³⁰ Indeed, economists have suggested that firms that operate in two-sided markets are more aptly called “two-sided platforms.”³¹ Rochet and Triole’s findings were quickly followed by an abundance of theoretical and empirical studies on the subject.³²

Although theoretical accounts of two-sided platforms are relatively new, two-sided platforms themselves are not new phenomena. David S. Evans, a leading economist on the subject, has identified many industries that simultaneously serve two distinct groups of consumers, including magazines (readers and advertisers), dating clubs (men and women), the stock exchange (buyers and sellers of securities), credit card networks (merchants and consumers), shopping centers (manufacturers and consumers), video game consoles (gamers and developers), operating systems (application developers and internet users), and search engine businesses (internet users and advertisers).³³ Although there is a lack of consensus on how to define two-sided platforms, it has been generally recognized that two-sided platforms are characterized by several unique features.

First, two-sided platforms cater simultaneously to “two distinct group of customers who need each other in some way and who rely on the platform to intermediate transactions between them.”³⁴ Second, there are indirect network effects between the customer groups of two-sided platforms. Therefore, a business strategy that affects one customer group necessarily affects the second customer group which, in turn, affects the first customer group and so on.³⁵ The existence of these indirect network effects raises “chicken and egg” issues, as a platform will not be able to attract customers on one side of a market if it does not have sufficient customers on the other side of the market.³⁶ Third, because the network effects among the customer groups can be very imbalanced, two-sided platforms can have a pricing structure skewed to balance the interdependent demands of consumers from both sides of the market.³⁷

In the case of Baidu, there are indirect network effects between internet users and advertisers, although such effects are very asymmetric. Companies that use

Baidu's search advertising services to promote their products value the service according to the number of users it attracts: More internet users means more exposure and a greater chance of reaching potential consumers. Such network effects can be inferred from Baidu's pricing structure, as Baidu's revenue is determined by the level of bidding fees and the amount of click-throughs to advertisers' websites.

On the other hand, it is not entirely clear whether internet users like or dislike the paid search results on Baidu's website. While those internet users specifically looking for advertising-related information might find Baidu less useful if there were fewer paid search results, others may care very little about paid search results. Moreover, in the case where natural search results are not clearly segregated from the paid search results (i.e. Baidu's previous practice), some internet users may find that the decrease of paid search results could even enhance their search experience.

Because the network effects among Baidu's customer groups were very imbalanced, Baidu adopted a pricing structure skewed to balance the interdependent demands of consumers from both sides of the market. The search engine service had been provided to users free of charge³⁸ and Baidu only charged advertisers for the marketing services. Indeed, according to its public filings, Baidu generated approximately 99.8 percent, 99.9 percent, and 99.9 percent of its total revenue from online marketing services in 2007, 2008, and 2009 respectively, and a substantial majority of this revenue was generated from the search advertising services.³⁹

DUE TO THE COMPLEXITY OF THE INTERRELATIONSHIPS AMONG CUSTOMERS GROUPS OF TWO-SIDED PLATFORMS, THE COMPETITION ANALYSIS OF SUCH PLATFORMS HAS PRESENTED SIGNIFICANT CHALLENGES TO ENFORCEMENT AGENCIES AROUND THE WORLD.

Due to the complexity of the interrelationships among customers groups of two-sided platforms, the competition analysis of such platforms has presented significant challenges to enforcement agencies around the world.⁴⁰

While there is no settled formula for assessing the market power of two-sided platforms, there is general agreement among competition authorities that accounting for the linkages between the two sides of the market is very important.⁴¹ Indeed, studies have shown that where indirect network effects are significant and are relevant for assessing the practices at issue, competition analyses that focus on one side of a business in isolation from the other side usually lead to errors.⁴² For instance, the fact that search engine providers set the price above marginal cost for the search advertising services does not really provide any useful indication of pricing to exploit market power. Otherwise, it would lead to the distorting conclusion that all search engine providers have market power over the search advertising services.

Because two-sided platforms have to coordinate demand between two interdependent customer groups, the analysis of market power must consider feedback

effects in order to determine the overall effect of a price change on profits.⁴³ Significantly, the capacity for two-sided platforms to exercise market power on one side can depend on the competitive restraints faced on the other.⁴⁴ Accordingly, a more sensible approach to evaluate Baidu's market power in this case would have been to adopt a two-sided approach that considered the feedback effects from both sides in order to determine the overall effect of a price change of the advertising services on the overall profits. As it is not entirely clear whether the feedback effects from the search engine side were positive or negative, rigorous economic analysis was needed to assess whether they were a meaningful constraint on Baidu's market power over the search advertising services.

Moreover, the difficulties of evaluating Baidu's market power are compounded by the limitations of the hypothetical monopolist test (also known as the "small significant non-transitory increase in price" test), a standard tool for defining the relevant market.⁴⁵ The first limitation is the well-known cellophane fallacy.⁴⁶ In cases involving abuse of dominant firm conduct, because the firm is already charging a supra-competitive price, the hypothetical monopolist test may apply too broadly by including products that would not have been close substitutes at a lower competitive price. China's guidelines on relevant market definition correctly recognizes this limitation: "[i]n such a circumstance, an adjustment to the competitive price is necessary for selecting a price that is more competitive."⁴⁷ In practice, however, it may be extremely difficult if not impossible to determine the competitive price level.⁴⁸ In fact, if one knows the competitive price, there is no need to define the market in the first place as it is obvious whether the current price exceeds the competitive price.⁴⁹

The second limitation, which is often overlooked, is that it can be misleading to apply the hypothetical monopolist test to high-tech companies because innovation can make it extremely difficult to identify substitute products.⁵⁰ Indeed, if one is to favor dynamic competition rather than static competition, a firm's monopoly power today may reveal very little about its future. As asserted by Teece and Coleman:

THE SECOND LIMITATION, WHICH IS OFTEN OVERLOOKED, IS THAT IT CAN BE MISLEADING TO APPLY THE HYPOTHETICAL MONOPOLIST TEST TO HIGH-TECH COMPANIES BECAUSE INNOVATION CAN MAKE IT EXTREMELY DIFFICULT TO IDENTIFY SUBSTITUTE PRODUCTS.

"Simply analyzing the market from a static perspective will almost always lead to the identification of markets that are too narrow. Because market power is often quite transitory, standard entry barrier analysis—with its 1-to 2-year fuse for entry—will often find that an innovator has power over price when its position is in fact extremely fragile."⁵¹

For example, Google, a dominant player in the search engine business, now faces a serious threat from new rivals such as Groupon, Facebook, Twitter, LinkedIn, and a few other companies that also provide platforms for online advertising. Analyzing Baidu's market power in this dynamic sense would have added an extra layer of complexity.

Given the above uncertainties, a court should be extremely cautious in applying the protocols of standard competition analysis to evaluate the market power of firms like Baidu. Indeed, while courts often begin the assessment of a monopolization case with the analysis of market definition, they should recognize that the assessment of the economic effects of alleged abuses is also an important screen for dismissing frivolous suits.⁵² If the pro-competitive effects of an abusive behavior clearly outweigh the exclusionary effects, or if the abusive behavior could be remedied quickly by competition, then there is no need to go through the strenuous exercise of evaluating the market power of a firm like Baidu.

B. INCENTIVES TO IMPOSE SWITCHING COSTS

Baidu claimed that it had blocked Renren's natural search results solely on account of the existence of junk links on Renren's website. According to the transcripts of the case, the Court did find many junk links on Renren's website and, moreover, Renren did not contest their existence. The Court therefore held that Baidu was justified in blocking Renren's website as doing so enhanced the reliability and accuracy of Baidu's search engine service.

However, the Court seemed to readily accept Baidu's defense without investigating further whether the blockage was solely motivated by the existence of junk links. Notably, prior to the adoption of the AML, the Court dealt with a similar case involving Baidu.⁵³ Beijing Land of Maple Travel and Cultural Exchange Ltd. ("LOM"), an operator of a travelling network (www.canada-travel.cn) began to use Baidu's paid search advertising services in July 2004. In March 2006, LOM noticed a sharp decline in its rankings in the natural search results of relevant key words. It then submitted a letter to Baidu complaining of unfair treatment. Baidu restored LOM's rankings in less than ten days.

Subsequently, LOM filed a suit against Baidu, alleging that Baidu had purposefully downgraded its natural search results because LOM had repeatedly refused requests from Baidu's to expand its usage of the search engine's paid search advertising services. LOM argued that Baidu's action was unfair competition and had infringed its reputation and property rights. The defense Baidu adopted was almost identical to the one later used in the Renren case—namely, that a large amount of junk links on LOM's website had caused Baidu's search engine system to automatically downgrade its natural search results. The Court in the LOM case was convinced by Baidu's reasoning and ultimately dismissed the case for lack of evidence.

A curious difference, however, is that Baidu restored LOM's rankings after receiving the letter of complaint, but it never did so in the *Renren* case. Baidu's contrasting reactions in these two cases reveal the enormous discretion it had in deciding how to penalize websites with junk links. Indeed, the information asymmetry between Baidu and customers such as LOM and Renren made it difficult to discern whether Baidu had downgraded the websites with the legitimate reason of penalizing junk links or with the motive of coercing those websites into using its advertising services.

Renren's lawyer argued that Baidu's behavior violated the exclusive dealing provision under Article 17(4). At first sight, this is not a typical exclusive dealing case. Baidu did not require Renren to deal exclusively with it, and Renren was free to use the search advertising services of other search engine providers. Nonetheless, it is possible to interpret the blocking of Renren's website as indication of a coercive scheme—one taking the form of an artificial switching cost imposed by Baidu on its existing customers. Under such a scheme, any customer who decides to stop using (or to use less of) the search advertising services risks losing the natural search results originally provided for free. Such a switching cost (if it in fact exists) can arise whenever an advertiser opts to use less service from Baidu, as in *Renren v. Baidu*: Renren didn't switch to another competing search engine provider, it simply reduced payment to Baidu for the search advertising services.

Artificial switching costs are not uncommon. They are sometimes purposely created by firms to lock in customers. The problem with such costs is that, after the initial purchase of a product, the consumer is locked in and cannot readily switch to another seller unless he or she is willing to pay the costs.⁵⁴ The existence of the switching costs thus makes it much harder for rivals to compete with Baidu: Even if such rivals could provide completely identical service, Baidu's existing customers would be reluctant to switch unless they were compensated for the loss of natural search results on Baidu's website. Accordingly, even though Baidu's practice fell short of outright exclusivity, it may still be treated as a weak form of exclusive dealing characterized by a high switching cost. Loyalty discounts (sometimes called "loyalty rebates"), for instance, have been argued to be weak forms of exclusive dealing because they are also characterized by high switching costs.⁵⁵

The reason Baidu has incentives to impose switching costs on its existing customers is closely related to its pricing structure. Baidu does not directly set the price for its advertising service; rather, it uses a generalized second-price auction system to sell the sponsored slots.⁵⁶ In a second-price auction, advertisers who pay the highest bid win the auction and pay the next highest bid. This may give

IT IS POSSIBLE TO INTERPRET THE BLOCKING OF RENREN'S WEBSITE AS INDICATION OF A COERCIVE SCHEME—ONE TAKING THE FORM OF AN ARTIFICIAL SWITCHING COST IMPOSED BY BAIDU ON ITS EXISTING CUSTOMERS.

incentives to Baidu to lock in existing customers in order to maximize the revenue derived from the search advertising services.

A simple example of a second-price auction can illustrate this point. Assume there are currently three bidders competing for the same key word: "Peking Duck." For the first slot, Bidder 1's willingness to pay is RMB 500; Bidder 2's willingness to pay is RMB 400; and Bidder 3's willingness to pay is RMB 300. For the second slot, Bidder 1's willingness to pay is RMB 400; Bidder 2's willingness to pay is RMB 300; and Bidder 3's willingness to pay is RMB 200. For the third slot, Bidder 1's willingness to pay is RMB 300; Bidder 2's willingness to pay is RMB 200; and Bidder 3's willingness to pay is RMB 100.

Assuming (a) each bidder is only interested in getting one slot for its website and (b) each bidder is interested in getting the highest possible slot, then Bidder 1 will be able to win the auction for the first slot with a bidding price at RMB 400 and Bidder 2 will be able to win the second slot with a price at RMB 300. As both Bidder 1 and Bidder 2 have received their highest possible slots, Bidder 3 has no competitors and only needs to bid above the minimum bidding price (let's assume it is zero for the simplicity of discussion).⁵⁷ Thus the final price for the first, second, and third slots will be RMB 400, RMB 300, and slightly above zero, respectively, and the auction generates approximately RMB 700 revenue.

Now, suppose Bidder 1 stops using the search advertising services, leaving only Bidder 2 and Bidder 3. The final price for the first and second slot would be RMB 300 and slightly above zero, respectively, ultimately generating revenue of approximately RMB 300. From this it can be seen that the loss of Bidder 1 would have ripple effects: Not only would Baidu lose the revenue for the payment for the second slot, but the revenue from the first slot would decrease as well.

C. REPUTATIONAL COSTS TO BAIDU

Assuming Baidu did try to coerce Renren into using its website by imposing an artificial switching cost, Baidu still faces a significant constraint—reputational costs. Economists have long considered the importance of reputation to be a private device that could help eliminate information asymmetries between buyers and sellers.⁵⁸ As noted by George A. Akerlof in his brilliant study on used car sales, brand name is an example of an institution that counteracts the effects of quality uncertainty by giving consumers the power to retaliate against firms by curtailing future purchases.⁵⁹ Indeed, Baidu is a leading search engine provider and one of most successful brand names in China.⁶⁰ Baidu's management would do well to remember the old Chinese saying "The water that bears the boat can also swallow it up"—that is, it should be well aware of the salient consequences of any potential damages to the firm's reputation.

To some extent, this case is similar to *Eastman Kodak v. Image Technical Services, Inc.* ("Eastman Kodak"),⁶¹ a case decided by the United States Supreme

Court in 1992. Kodak manufactures and sells high-volume photocopiers and micrographic equipment. At the time of the sale, Kodak sold repair parts, enabling users to repair their copiers or hire independent service organizations (“ISOs”) to do so. Kodak later changed its policy of supplying repair parts to ISOs and confined sales of repair parts to Kodak copier owners who contracted to have their copiers serviced by Kodak. The ISOs then accused Kodak of unlawfully tying the sale of service to the sale of parts and unlawfully monopolizing and attempting to monopolize the sale of service.

Since Kodak had little market share in the original equipment market, the court held that no unlawful tie could exist between Kodak original equipment and Kodak parts-servicing. The court focused instead on whether an unlawful tie-in existed between Kodak parts and Kodak servicing. The court conceded that customers who had anticipated the change of policy could not have been exploited; they would have shopped around and purchased copiers based on full lifecycle costs. However, the court found that, because the information required was so difficult and costly to come by, a substantial number of consumers did not enjoy cost-efficient access to the pricing information needed to evaluate such life-cycle cost. Moreover, the court found that a current Kodak-copier owner might tolerate even uncompetitive price increases in Kodak parts and services so long as the increases did not exceed the cost of abandoning his or her original investment in a Kodak copier and switching to another copier. For these reasons, the court denied Kodak’s motion for summary judgment.

This case sparked tremendous controversy among economists and legal practitioners.⁶² Dennis W. Carlton, for example, criticized the decision for the failure to recognize that *ex ante* competition completely protects consumers.⁶³ Judge Posner also questioned the decision—even if Kodak had monopoly power in the original equipment market, the reputational cost to Kodak would have been likely to deter any exploitative behavior that would make new entry into the market attractive.⁶⁴ Some economists, on the other hand, pointed out that declining profits in the equipment market or a significant base of equipment owners may induce a profit-maximizing firm to engage in *ex post* exploitation of consumers, also called installed-based opportunism.⁶⁵ However, such installed-base opportunism is less attractive to any firm that has a desire to increase its market size—the loss of future profits due to the damage to its reputation would outweigh the gains from short-term profits derived through exploiting the aftermarket.⁶⁶

As in *Eastman Kodak*, if Baidu imposes (or, for that matter, continues to impose) an artificial switching cost on its existing customers, it may well enjoy increased market power over its existing customers. However, with the exposure of such abusive behavior come fully informed customers. Existing customers who are locked in would be unlikely to choose Baidu again and new customers would be less willing to choose Baidu for fear that they would be locked in later. This would make new entry into the market more likely.

Further, the reputational harm to Baidu would mean that it would need to compete more aggressively for new customers and, thus, *ex ante* competition would benefit new customers. When price discrimination is possible, firms will price more aggressively in the first period to attract new customers, as the first-period market share (or customer base) has a positive impact on the second-period profit.⁶⁷ However, in this case, as the search advertising services is auction-based, Baidu does not have direct control over its price. In order to attract more new customers, Baidu needs to provide better marketing services to advertisers and better search engine service to users (which, in turn, makes marketing services more attractive to advertisers). Overall, it is not clear whether the net effect on competitiveness due to switching costs would result in higher or lower total consumer welfare.

FURTHER, THE REPUTATIONAL HARM TO BAIDU WOULD MEAN THAT IT WOULD NEED TO COMPETE MORE AGGRESSIVELY FOR NEW CUSTOMERS AND, THUS, EX ANTE COMPETITION WOULD BENEFIT NEW CUSTOMERS.

Overall, it is not clear whether the net effect on competitiveness due to switching costs would result in higher or lower total consumer welfare.

More importantly, as Baidu has experienced a tremendous amount of growth in the past few years,⁶⁸ any fly-by-night strategy to gain temporary profits (such as the alleged abusive behavior) would not be profit maximizing for Baidu.

In a recent interview with the *Wall Street Journal*, Baidu's chief executive officer Robin Li emphasized that he ran the company based on a vision of long-term growth rather than one of short-term investor expectations.⁶⁹ Mr. Li also touted Baidu's tremendous growth potential, noting that two-thirds of the Chinese population has yet to learn how to use the internet.⁷⁰ This reveals that the management of Baidu may have little incentive to gain short-term profits at the expense of future growth. In fact, since the media exposed the *Renren v. Baidu* case, Baidu immediately posted a statement on its website declaring that it had never and would never use any coercive measures to force companies to use its promotional service. Baidu claimed, furthermore, that it would investigate and penalize any salesperson using such methods to promote sales.⁷¹

From this perspective, consumer protection law, or even widespread media disclosure of a scandal that could jeopardize Baidu's reputation, would be sufficient to deter the alleged abusive behavior. Since informed consumers would not be harmed by such abusive behavior, the application of the AML to this case is like using a sledgehammer to crack a nut. Even if Baidu is found to have market power, the free market corrects such alleged abusive behavior much faster than antitrust law does.

V. Conclusion

As the first abuse of a dominant position case involving a two-sided platform in China, *Renren v. Baidu* presented significant challenges to the Court, which had almost no precedent to rely on when deciding it. Although the Court recognized certain two-sided features of Baidu's business model, it failed to explore further

the impact of those features on competition analysis. Crucially, the Court erred in defining the relevant product market as the search engine service market. Instead of using a one-sided approach, the Court should have adopted a two-sided approach in defining the relevant market. Moreover, the problems confronted in attempting to identify the relevant market in this case were compounded by the difficulties entailed in applying the hypothetical monopolist test, which is vulnerable to the cellophane fallacy. Finally, the hypothetical monopolist test does not function very well when analyzing the relevant market for high-tech companies as innovation can make it extremely difficult to identify substitute products.

Given the complexities of identifying the relevant market and assessing the market power of Baidu, the evaluation of the economic effects of allegedly abusive behavior becomes a very important screen for dismissing frivolous lawsuits. Although the Court did find the existence of junk links on Renren's website, the information asymmetry between Baidu and customers such as Renren made it difficult to discern whether Baidu had downgraded the websites with the legitimate reason of penalizing junk links or with the motive of coercing those websites into using its advertising services. As Baidu's business model is based on an auction system, there is a theoretical possibility that Baidu may have an incentive to impose artificial switching costs in order to lock in existing customers. The existence of the switching costs could raise rivals' costs as Baidu's existing customers would be reluctant to switch unless they were compensated for the loss of natural search results on Baidu's website.

However, the reputational costs to Baidu should be sufficient to deter such alleged abusive behavior. New customers who are informed about the switching cost would be unlikely to choose Baidu, and existing customers who are locked in would be unlikely to choose Baidu again. Besides, Baidu would need to compete more aggressively for new customers by improving its search engine and search advertising services. Therefore, it is not clear whether the net effect on competitiveness due to switching costs would result in higher or lower total consumer welfare. More importantly, as Baidu expects to have a great potential for future growth, it would not find it profit maximizing to adopt a fly-by-night strategy to gain temporary profits at the expense of future growth. As informed consumers would not be harmed, consumer protection law rather than antitrust law would have been a better tool for tackling abuses like those alleged by Renren in this case. ▼

GIVEN THE COMPLEXITIES OF IDENTIFYING THE RELEVANT MARKET AND ASSESSING THE MARKET POWER OF BAIDU, THE EVALUATION OF THE ECONOMIC EFFECTS OF ALLEGEDLY ABUSIVE BEHAVIOR BECOMES A VERY IMPORTANT SCREEN FOR DISMISSING FRIVOLOUS LAWSUITS.

1 See e.g., Xinzhu Zhang & Vanessa Zhang, *Chinese Merger Control: Pattern and Implications*, 6 (2) J. COMPETITION L. & ECON.477(2010); Huyue (Angela) Zhang, *Problems in Following EU Competition Law*:

- A Case Study of Coca-Cola/Huyuan*, PEKING U. J. LEGAL STUDIES (forthcoming), available at: <http://ssrn.com/abstract=1569836>
- 2 For a discussion on the latest development of China's antitrust enforcement and the problems in the enforcement system, see Huyue (Angela) Zhang, *The Enforcement of China's Anti-Monopoly Law, An Institutional Design Perspective*, ANTITRUST BULL. (forthcoming), available at: <http://ssrn.com/abstract=1783037>.
 - 3 See Baidu. Inc.'s annual report for the year ended December 31, 2009.
 - 4 See Opinion by Beijing No. 1 Intermediate People's Court, Civil Case No. Yizhongminchuzi 845/2009. The transcripts of the case are available (in Chinese) at: http://www.chinacourt.org/zhibo/zhibo.php?zhibo_id=1865.
 - 5 See, e.g., Dr. R. Ian McEwan, *China- the Baidu Decision*, COMPETITION POLICY INT'L 6(2) (2010); Fei Deng, Dr. Gregory K. Leonard, & Wenfeng Tang, *How Private Antitrust Litigation May Proceed in China*, LAW 360, January 5, 2010; Cleary Gottlieb Steen & Hamilton, ASIAN COMPETITION REPORT, October to December 2009, available at www.cgsh.com.
 - 6 *Id.*
 - 7 See Baidu Inc.'s annual report, *supra* note 3.
 - 8 *Id.*
 - 9 *Id.*
 - 10 Baidu's pricing information is available at its promotional website (in Chinese): <http://e.baidu.com/fwjs/cjwvt/2009-01-04/170548209216.html>.
 - 11 See Baidu Inc.'s annual report, *supra* note 3.
 - 12 See Daniel F. Spulber, *The Map of Commerce: Internet Search, Competition, and the Circular Flow of Information*, 5(4) J. COMPETITION L. & ECON. 633(2009).
 - 13 See Letter from Heather Hipsley, Acting Assoc. Dir., Div. of Adver. Practices, FTC to Gary Ruskin, Executive Dir., Commercial Alert (June 27, 2002), available at <http://www.ftc.gov/os/closings/staff/commercialalertletter.shtml>
 - 14 After the exposure to several advertising scandals, Baidu attempted to recover its reputation by introducing the Phoenix Nest system, which provides a better distinction between paid search results and natural search results. More information is available at Baidu's official website: <http://e.baidu.com/>. Note the dispute between Renren and Baidu occurred prior to the introduction of this new system.
 - 15 See Ming Shuliang & Zhang Poling, *Baidu: the Transmutation of the Bid for Ranking System* (in Chinese), CAIJING, Issue 5, December 8, 2008.
 - 16 *Id.*
 - 17 See generally Viva R. Moffat, *Regulating Search*, 22(2) HARV. J. L. & TECH. 475(2009) (tracing the contours of the academic debate on search engine regulation); See also Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1178-88 (2008) (arguing that a Federal Search Commission could be created to regulate search engines); Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J. L. & TECH. 188, 195-198 (2006) (advocating a free market approach in the regulation of search engines).

- 18 Unless otherwise specified, the information discussed in this section is based on the transcripts of this case, see *supra* note 4.
- 19 See Tong Shu, *Reflections on Baidu Monopoly Litigation: Comments on Renren v. Baidu*, 1 CHINA PATENTS & TRADEMARKS 66 (2010).
- 20 According to Renren, Baidu posted four results of its website search because Renren was still using the search advertising services. So it appears that those four results were paid search results and that Renren's natural search results had all been blocked.
- 21 2007 WL 831806 (N. D. Cal.).
- 22 Shu, *supra* note 19.
- 23 2007 WL 831806, at *1.
- 24 *Id.*, at *4-12.
- 25 *Id.*, at *5-6.
- 26 *Id.*, at *5.
- 27 See Shu, *supra* note 19.
- 28 See William F. Baxter, *Bank Exchange of Transactional Paper: Legal and Economic Perspective*, 26 J. L. & Econ. 541 (1983).
- 29 See Jaen-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Market*, 1 J. EUR. ECON. ASS'N 990 (2003).
- 30 See Organization for Economic Co-operation and Development ("OECD"), *Two-Sided Markets*, DAF/COMP (2009) 20, 28.
- 31 *Id.*, at 11.
- 32 See e.g. Mark Armstrong, *Competition in Two-Sided Markets*, 37(3) RAND J. ECON., 668(2006); Bernard Caillaud & Bruno Jullien, *Chicken and Egg: Competition among Intermediation Service Providers*, 34(2) RAND J. ECON., 521(2003); David S. Evans, *The Antitrust Economics of Multi-Sided Platform Markets*, 20 YALE J. ON REG., 325(2003); Geoffrey Parker & Marshall Van Alstyne, *Two-Sided Network Effects: A Theory of Information Product Design*, 51(10) MGMT. SCI. 1494(2005).
- 33 See Evans, *id.*
- 34 *Id.*
- 35 See David S. Evans & Michael D. Noel, *The Analysis of Mergers that Involve Multi-Sided Platform Businesses*, 4(3) J. COMPETITION L. & ECON. 663(2008).
- 36 See Rochet & Tirole, *supra* note 29.
- 37 See OECD, *supra* note 30, at 11-12.
- 38 As economists often put it, there is no such thing as a free lunch. Although the search engine service has been provided to internet users free of charge, users pay for the service by revealing personal information about themselves through their online activities, viewing advertisements, and ultimately patronizing online advertisers. See Spulber, *supra* note 12.

- 39 See Baidu's Inc.'s annual report, *supra* note 3. Other than search advertising, Baidu also provides other forms of online marketing services such as graphical advertisements.
- 40 See OECD, *supra* note 30, at 11.
- 41 *Id.*
- 42 See Evans & Noel, *supra* note 35.
- 43 See OECD, *supra* note 30, at 24.
- 44 *Id.*, at 79.
- 45 Although the Court in this case seemed to have considered the substitutability of the search engine service with other types of internet service, it is not entirely clear whether the Court had applied the hypothetical monopolist test in assessing the relevant market.
- 46 The concept of the cellphone fallacy is derived from *United States v. E.I. du Pont de Nemours & Co.*, 351 U.S. 377 (1956), in which the court defined the market too broadly by using the current market price (which was above the competitive level) to assess the cellophane producer's market power.
- 47 See Article 11 of the Guidelines on Relevant Market Definition promulgated by the Anti-Monopoly Committee of the State Council on May 24, 2009.
- 48 See Office of Fair Trading, *The Role of Market Definition in Monopoly and Dominance Inquiries*, Economic Discussion Paper 2, July 2001; see also American Bar Associations, *Joint Comments of the American Bar Association Section of Antitrust Law and Section of International Law of the MOFCOM Draft Guidelines for Definition of Relevant Market*, January 30, 2009.
- 49 See Dennis W. Carlton, *Does Antitrust Need to Be Modernized?* 21 J. ECON. PERSPECTIVE 164 (2007).
- 50 See e.g., J. Gregory Sidak & David J. Teece, *Dynamic Competition in Antitrust Law*, 5(4) J. COMPETITION L. & ECON 581 (2009); Christopher Pleatsikas & David Teece, *The Analysis of Market Definition and Market Power in the Context of Rapid Innovation*, 19 INT'L J. INDUS. ORG. 665 (2001); David J. Teece & Mary Coleman, *The Meaning of Monopoly: Antitrust Analysis in High-Technology Industries*, 43 ANTITRUST BULL. 801 (1998); Raymond Hartman, David Teece, Will Mitchell, & Thomas Jorde, *Assessing Market Power in Regimes of Rapid Technological Change*, 2(1) INDUS. & CORP. CHANGE 317 (1993).
- 51 See Teece & Coleman, *supra* note 50.
- 52 See Carlton, *supra* note 49.
- 53 See Opinion by Beijing No. 1 Intermediate People's Court, Civil Case No. Yizhongminchuzi 12408/2006.
- 54 See Alan Beggs & Paul Klempere, *Multi-Period Competition with Switching Costs*, 60(3) ECONOMETRICA, 651-666 (1992).
- 55 ROBERT O' DONOGHUE & A. JORGE PADILLA, *THE LAW AND ECONOMICS OF ARTICLE 82 EC*, 374 (2006).
- 56 Baidu's pricing model is based on a generalized second price auction system but it also takes into account the quality factor of the key words, thus its pricing is not as transparent as a standard generalized second price auction. However, this does not affect the analysis here. See Michael Ostrovsky, Benjamin G. Edelman, & Michael Schwarz, *Internet Advertising and the Generalized Second Price*

- Auction: Selling Billions of Dollars Worth of Keywords*, 97(1) AM. ECON. REV. 242 (2007) (explaining the mechanism of generalized second price auction); see also Renato Gomes, *Sponsored Search Auctions: Simple Economics and Implications for Antitrust Policy*, 6(2) COMPETITION POL'Y INT'L (2010).
- 57 Since June 2006, Baidu introduced a dynamic mechanism for the determination of the minimum bidding price for each keyword. See Baidu, Inc.'s annual report for the year ended December 31, 2007.
- 58 See e.g., Friedrich A. Hayek, *The Meaning of Competition*, in INDIVIDUALISM AND ECONOMIC ORDER (1948).
- 59 See George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 QUART. J. ECON. 488 (1970); see also Benjamin Klein & Keith B. Leffler, *The Role of Market Forces in Assuring Contractual Performance*, 89 J. POL. ECON. 615 (1981).
- 60 See Jez Frampton, *China: Brand is A Missed Opportunity*, Forbes.com, July 28, 2010, available at: http://www.forbes.com/2010/07/28/china-marketing-brands-baidu-haier-lenovo-cmo-network-inter-brand_print.html; see also Interbrand and Forbes Release Best Chinese Brand 2010, July 6, 2010, available at http://www.interbrand.com/press_release.aspx?pressid=328&langid=1000 (According to this article, Baidu was ranked as the 17th most valuable brand in China and the compound annual growth rate of its brand was 85 percent in 2010).
- 61 See *Eastman Kodak Co. v. Image Technical Servs., Inc.*, 504 U.S. 451 (1992).
- 62 See e.g., Carl Shapiro, *Aftermarkets and Consumer Welfare: Making Sense of Kodak*, 63 ANTITRUST L.J. 483 (1995); Severin Borenstein, Jeffery K. MacKie-Mason, & Janet S. Netz, *Antitrust Policy in Aftermarkets*, 63 ANTITRUST L. J. 655 (1995); Dennis W. Carlton, *A General Analysis of Exclusionary Conduct and Refusal to Deal—Why Aspen and Kodak Are Misguided*, 68 ANTITRUST L. J. 659 (2001).
- 63 See Carlton, *id.*
- 64 See RICHARD A. POSNER, ANTITRUST LAW 237 (2d ed. 2001).
- 65 See Jeffery K. MacKie-Mason & Janet S. Netz, *Antitrust Policy in Aftermarkets*, 63 ANTITRUST L. J. 655 (1995).
- 66 See Carl Shapiro, *Aftermarkets and Consumer Welfare: Making Sense of Kodak*, 63 ANTITRUST L.J. 483 (1995).
- 67 MASSIMO MOTTA, COMPETITION POLICY: THEORY AND PRACTICE 80 (2004).
- 68 Baidu's online marketing revenue increased by 83.5 percent and 39.2 percent in 2008 and 2009, respectively. See Baidu Inc.'s annual report for the year ended December 31, 2008 and 2009, respectively. Baidu's performance was also remarkable in 2010, with online marketing revenue increased by 59.6 percent and 74.5 percent in the first and second quarter, respectively. See Baidu Inc.'s Form 6-K filed on April 29, 2010 and July 23, 2010, respectively.
- 69 Owen Fletcher, *Baidu's CEO Pursues Long-Term Growth*, WALL ST. J., August 4, 2010, B8. ("I don't run the company based on investor expectations. I run the company based on our own vision of the future of Internet computing and the future of the Chinese market. I'm the founder. I will stay here for a very long time. I don't need to please those short term investors for next quarter. I need to make sure the company is healthy and strong and will continue to grow for many, many years.")
- 70 *Id.*
- 71 Baidu's statement (in Chinese) is available at <http://e.baidu.com/fwjs/cjwtf/>.



CPI COMPETITION POLICY
INTERNATIONAL

Volume 7 | Number 1 | Spring 2011

Jeffrey Rohlfs' 1974 Model of
Facebook: An Introduction
with
*A Theory of Interdependent
Demand for a Communications
Service*
by Jeffrey Rohlfs

Richard Schmalensee

Massachusetts Institute of
Technology

Jeffrey Rohlfs' 1974 Model of Facebook: An Introduction

*Richard Schmalensee**

I. Introduction

Jeffrey Rohlfs' pioneering 1974 study of demand in the presence of network externalities,¹ which make each actor's demand for some good or service depend in part on whether others purchase it, laid the foundation for a huge academic literature that has had a major impact on antitrust policy. The government's case in *U.S. v. Microsoft*, for instance, relied heavily on network externality arguments.

In most of the post-Rohlfs network-effects literature, buyers are modeled as making long-term product or technology choices because those choices either involve the purchase of significant durable goods or create switching costs. Examples include the choices between VHS and Betamax VCRs, or between Apple and Wintel computers, or the choice to purchase an early fax machine.

In contrast, Rohlfs presents a model that seems better suited to analysis of new Internet-based businesses that rely on network effects, like Facebook and YouTube. These businesses provide services rather than durable goods, and their customers are not required to make long-term commitments. Switching costs are at most moderate, and customers can often participate in multiple competing networks at the same time. (In the terminology of the recent, related literature on two-sided markets, they can “multi-home.”²) Thus I think the Rohlfs paper deserves to be read carefully on its own, apart from the literature it helped to launch.

*Howard W. Johnson Professor of Economics and Management, Massachusetts Institute of Technology. I am indebted to David Evans and Joe Farrell for helpful comments, but errors and opinions are mine alone. It is a particular pleasure to help increase awareness of the Rohlfs paper both because its insights have shaped some of my own recent work (as discussed below) and because I remember Jeff fondly from our time together in graduate school at MIT.

Rohlf's fundamental assumption was that the amount any individual firm or household would be willing to pay for a telecommunications service would depend on the set of other individuals with whom they could use that service to communicate. This effect, what we now call a *direct* network effect or externality, seems to have been recognized from the first days of the telephone industry. (The recent two-sided markets literature, in contrast, focuses mainly on *indirect* network externalities, in which participation by members of one customer group makes a platform more attractive to members of another customer group.) Rohlf's was not the first formal analysis to incorporate this assumption: his paper cites earlier articles by Artle & Averous and by Squire that do so.³ Those papers were concerned with optimal telecommunications pricing, however, while Rohlf's generally took price as given and provided a deeper, more general analysis of the implications of network externalities for market demand—without, it should be noted, ever using the term “network externalities.”

ROHLF'S GENERALLY TOOK PRICE AS GIVEN AND PROVIDED A DEEPER, MORE GENERAL ANALYSIS OF THE IMPLICATIONS OF NETWORK EXTERNALITIES FOR MARKET DEMAND—WITHOUT, IT SHOULD BE NOTED, EVER USING THE TERM “NETWORK EXTERNALITIES.”

The next section provides a guide to Rohlf's analysis. It is, I hope, a bit easier to digest than the paper it attempts to exposit, but it is intended mainly for economists and lawyers who are tolerant of formal reasoning. None of its equations are essential for Section III, which describes the impact Rohlf's paper has had and considers some of its implications for both economic analysis of and competition policy toward Facebook-like businesses.

II. A New Telecommunications Service

Writing as a Member of the Technical Staff at Bell Laboratories, Rohlf was mainly concerned with the problem of launching a new telecommunications service (his main example, in 1974, was a video communications service) and thus with disequilibrium situations. His most general model (in Section 2) begins with the assumption that, all else equal, if any individual i subscribes to the telecommunications service under consideration, her utility will not be decreased and may be increased if any other individual j also subscribes. Since in the 1970s customers of the Bell System could only rent telephones and other terminal equipment, it was natural for Rohlf to assume that subscriptions to a new service would involve only a per-period price, with no fixed cost of subscribing or unsubscribing.

Under these assumptions, Rohlf defines an *equilibrium user set*: taking prices and all other individuals' status as given, each individual in such a set wishes to continue to subscribe to the service, and each individual not in the set does not wish to subscribe. This is a natural Nash equilibrium concept, but note that it

rules out both explicitly coordinated behavior and non-myopic behavior that depends on expectations of others' future actions. Since decisions to subscribe or unsubscribe could be reversed at little, if any, cost in the regulated telecommunications sector circa 1974, it is not unreasonable to assume that expectations about the future would not affect an individual's current actions. As we discuss below, however, coordination devices could nonetheless be both privately and socially valuable.

Rohlfs notes that it is a fundamental feature of this model that equilibrium user sets are not generally unique. Consider, for example, a population of four individuals in which individuals 1 and 2 are willing to pay a lot to be able to communicate with each other but neither has any interest in individual 3 or individual 4. Suppose the situation is symmetric, so that individuals 3 and 4 are eager to communicate with each other but care nothing for 1 or 2. Then, if the service is of no value unless it enables you to communicate with somebody you care about, and if its price is low enough, there will be four equilibrium user sets: the null set, individuals 1 and 2, individuals 3 and 4, and all four individuals.

Whenever the service is of no value to any individual unless at least one other individual subscribes, the null set, in which no one subscribes, is always an equilibrium user set. Similarly, in more general settings there may also be a dramatic difference between the largest and smallest equilibrium user sets, and, as Rohlfs

REAL MARKETS WITH LARGE
NUMBERS OF PARTICIPANTS RARELY
LEAP INSTANTLY TO EQUILIBRIUM,
PARTICULARLY WHEN THEY INVOLVE
NOVEL GOODS OR SERVICES.

notes, "In a practical situation, this difference may mean the difference between marketing success and failure."⁴

Real markets with large numbers of participants rarely leap instantly to equilibrium, particularly when they involve novel goods or services. To reflect this, Rohlfs considers a

broad class of adjustment processes in which, out of equilibrium, each individual who wishes to subscribe or unsubscribe does so with some finite lag. In general, the equilibrium set to which such processes converge depends on the initial set of subscribers and, possibly, on the details of the adjustment process. In particular, Rohlfs notes, "It may be critical whether or not the disequilibrium non-users subscribe before the disequilibrium users drop out."⁵

In order to obtain sharper results, Rohlfs turns in his Section 3 to the special case in which the utility of the service to any one individual depends only on the number of other individuals who also subscribe, which he refers to as the *uniform calling* model, and the relationship is linear. This portion of the paper has perhaps been the most influential. Let f be the fraction of the total relevant population that subscribes. Then for a given positive price, p , a typical individual i will want to subscribe to the service if and only if

$$fw_i - p \geq 0, \text{ or } w_i \geq p/f. \tag{1}$$

where w_i is a non-negative constant. For a large population, one can treat the w_i as distributed according to a smooth probability density function, $h(w)$.

Given all the assumptions in the preceding paragraph, the shape of $h(w)$ and the corresponding distribution function, $H(w)$, will determine the characteristics of market demand. Taking price as given, if the fraction of the population currently subscribing is f , the fraction of the population who will want to subscribe once they know this, f^* , is just the fraction of the population for which equation (1) is satisfied:

$$f^*(f|p) = 1 - H(p/f). \quad (2)$$

Because distribution functions are non-decreasing, the function $f^*(f)$ is non-increasing in p and non-decreasing in f . If $f^*(f_i|p) = f_i$, then, given p , there is an equilibrium user set consisting of the fraction f_i of the population with the largest values of w . As long as price is positive, the null set ($f = 0$) is always an equilibrium user set, since $H(\infty) = 1$ for any distribution function. There may be no equilibrium user sets with $f > 0$, or there may be one or more such sets. If $H(p) = 1$, so that everyone in the population is willing to pay price p , then $f^*(1|p) = 1$, and the whole population is an equilibrium user set.

To analyze disequilibrium situations, Rohlfs considers a general class of adjustment mechanisms according to which f increases over time if $f < f^*$, and f decreases over time if $f > f^*$. It is then easy to show that if there are equilibria with $f > 0$, stable and unstable equilibria must alternate.⁶ That is, if the null set is a stable (unstable) equilibrium, the next smallest, if it exists, must be unstable (stable), and so on. Rohlfs defines the *critical mass* problem for the new system as the problem of somehow reaching a level of f such that $f^* > f$ and the business is viable at the next highest stable equilibrium, to which the system will then tend over time if price remains fixed. He notes, however, that attaining the socially optimal equilibrium user set “may require ruinous (albeit temporary) promotional costs.”⁷ Moreover, he adds that although he naturally assumes in his analysis “that the product is viable, it is worth noting that in real life the seller would have no such guarantee.”⁸

If there are multiple stable equilibria, it may be advantageous for both the seller and its customers (for whom participation by others adds value) if the market attains an equilibrium with high participation rather than a lower equilibrium or even the null equilibrium in which there is zero participation. For this reason, devices to coordinate behavior may have both private and social value. Thus, for instance, Glen Weyl analyzes “insulating tariffs,” in which prices are carefully set as functions of participation levels so that the market is guided to the desired equilibrium.⁹ In practice, of course, particularly with highly innovative new products, the distribution of reservation prices is unknown and thus so are the available equi-

DEVICES TO COORDINATE
BEHAVIOR MAY HAVE BOTH
PRIVATE AND SOCIAL VALUE.

libria. And, partly as a consequence, investors' enthusiasm for spending on costly coordination attempts is typically well curbed. On the other hand, it is not unusual for new products with network effects to raise price over time as participation grows and the product thus becomes more attractive,¹⁰ and Rohlfs considers strategies of this sort. One can think of these as developing and announcing pieces of insulating tariffs on the fly.

Much of the analysis in Rohlfs' Section 3 is devoted to an example in which w is distributed uniformly between zero and one, so that $H(w) = w$ over that interval. Then (2) becomes simply:

$$f^*(p|w) = \begin{cases} 1 - (p|f), & \text{for } 0 < p < f \leq 1. \\ 0, & \text{for } p > f. \end{cases} \quad (3)$$

Solving equation (3) for $f = f^*$, there are at most three distinct equilibria, which are shown in Figure 1:

$$f_0 = 0, f_1 = \frac{1 - \sqrt{1 - 4p}}{2}, \text{ and } f_2 = \frac{1 + \sqrt{1 + 4p}}{2}. \quad (4)$$

Note that if $p > 1/4$, the only equilibrium involves no subscribers, even though if everybody subscribed, 3/4 of them would gladly pay more than 1/4. The problem is that those willing to pay less than 1/4 even with the entire population subscribing would leave the service, reducing its attractiveness to those who remained, inducing further defections, and so on until the business spiraled down to zero.

The arrows in Figure 1 illustrate the dynamics of this example for some $p < 1/4$. The smallest equilibrium, f_0 , is stable; the next smallest, f_1 , is unstable, and the largest, f_2 , is stable. At the given price, the critical mass problem is to get the fraction of the population subscribing to above f_1 . If that can't be done, the subscriber base will inevitably shrink to zero, but if it is done, network effects will fuel organic growth to the largest equilibrium, f_2 .

In Section 4, Rohlfs generalizes the linear utility model in equation (1) to a situation in which the population can be divided into k groups. Initially he assumes that individuals in group i care only about the number of other individuals in group i who subscribe; then he allows for

ROHLFS' SECTION 5 CONSIDERS
TWO APPROACHES TO
SOLVING THE START-UP PROBLEM
FOR A NEW SERVICE.

inter-group externalities. In the latter case the analysis involves considerations of group-specific critical mass levels, and the ultimate equilibrium reached from any specific starting point depends on the details of both that point and the adjustment process.

Rohlfs' Section 5 considers two approaches to solving the start-up problem for a new service: 1) free service to a carefully selected group of people for a limited period of time, and 2) a low introductory price that is raised over time. If being

a subscriber has value only to the extent that it enables communication (rather than, say, because it is a status symbol), the hard part is getting two or more individuals with high values of w to subscribe at a positive price. In theory, at least, this might be done by offering the service for free for a limited time to a targeted group, then raising price just above zero so that at least a few high- w individuals find it optimal to remain subscribers.

Once this has been done successfully, however, at least in the uniform calling case, price can then be increased gradually over time according to some function $p^g(f)$. As long as $p^g(f) < f(1 - f)$, f^* will exceed f from (3), and the subscriber base will grow. (Once again, it doesn't matter whether individuals anticipate later price increases or not, since myopic behavior is individually rational here.) When the optimal level of price is reached, price is constant thereafter, and the system adjusts to the higher, stable equilibrium corresponding to that price. These pricing policies are, as I noted above, broadly in the spirit of Weyl's insulating tariffs, but Rohlfs seems to view them as generally being developed as information about demand arrives, rather than as the results of *ex ante* optimization with demand known.

Rohlfs establishes a very neat result in this context. Suppose that at some point, after the price has been raised to $p_0 > 0$ and all those who find it optimal to unsubscribe have done so, a fraction $f_0 > 0$ of the population finds it optimal to remain on the system. Then as long as $p^g(f)$ is less than $f(1 - f)$, as above, and the elasticity of $p^g(f)$ with respect to its argument does not exceed one, the system will grow and no subscriber will ever leave.¹¹

All this rests on the extremely unrealistic assumption that the distribution of the w_i is somehow known, of course. And, as Rohlfs shows, if the uniform calling assumption does not hold, devising a startup strategy necessarily becomes more complex even if all taste distributions are known, and the details of non-uniformity matter.

III. Impact and Implications

Jeffrey Rohlfs' 1974 paper has been widely cited—669 times according to Google Scholar—but its importance beyond telecommunications took some time to be recognized.¹² The first widely cited paper that I can find that cites the Rohlfs paper is a 1980 survey of the economic theory of clubs.¹³ The literature on network externalities, and with it citations of the Rohlfs paper, exploded in the mid-1980s with the publication of influential papers by Joseph Farrell and Garth Saloner and by Michael Katz and Carl Shapiro that acknowledged Rohlfs' contribution.¹⁴

As noted above, however, the literature that grew out of these papers deals with very different market environments than those considered by Rohlfs. In a widely cited 1994 survey, Michael Katz and Carl Shapiro note that in this literature's

analysis of technology adoption and product selection decisions, expectations and coordination play important roles.¹⁵ As I noted above, expectations and coordination do not appear at all in the Rohlfs analysis, with the exception that he considers increasing price paths as a device to attain a desirable equilibrium.

Katz and Shapiro's initial example of a technology adoption decision is the decision to buy a fax machine.¹⁶ In that decision expectations about the behavior of other potential fax machine owners are clearly important. Had fax machines existed in the early 1970s when Rohlfs was writing, however, they would have been available only for rent from the local telephone company, and expectations and coordination would have been of much less importance in decision-making about subscribing to fax services. Similarly, expectations and coordination were clearly important in making many of the product selection decisions between incompatible rival "hardware/software" systems that they describe as being "in the newspaper almost every day."¹⁷ These include the choice

between Beta and VHS video recording systems and among rival home video game systems. Again, if short-term rental of the "hardware" parts of these systems were available, expectations and coordination would have been much less important.

The focus of this post-Rohlfs network-effects literature on technology and product selection decisions with long-lasting consequences sent two important messages to competition policy-makers.

The first was implicit: technology adoption and product selection were generally modeled as discrete, once-and-for-all decisions that typically produced winner-take-all results. Not only is that how the choice between Beta and VHS seemed to most observers at the time, but to model multi-stage processes, in which expectations of future technologies and products (like DVDs and Blu-Ray and ...) could influence today's choices, would have involved considerable incremental complexity. The second was explicit: most theoretical analysis showed that market outcomes in markets with once-and-for-all competition and network effects could be seriously socially inefficient: buyers could find themselves selecting the wrong product or technology, and society could be locked-in to those bad choices for the foreseeable future.

Katz and Shapiro do counsel caution in their 1994 survey and note that, "In short, we are far from having a general theory of when government intervention is preferable to an unregulated market outcome."¹⁸ The literature they surveyed nonetheless suggested that, at the very least, competition authorities should pay particular attention to competition in industries with network effects to ensure that firms with short-run market power don't use anticompetitive behavior and network effects to build and lock in dominant positions for the long haul. As Carl

THE FOCUS OF THIS POST-ROHLFS
NETWORK-EFFECTS LITERATURE ON
TECHNOLOGY AND PRODUCT
SELECTION DECISIONS WITH LONG-
LASTING CONSEQUENCES SENT TWO
IMPORTANT MESSAGES TO
COMPETITION POLICY-MAKERS.

Shapiro, then Deputy Assistant Attorney General for Economics in the Antitrust Division of the U.S. Department of Justice, put it in a January, 1996 speech:¹⁹

“Even more so than in other areas, antitrust policy in network industries must pay careful attention to firms’ business strategies, the motives behind these strategies, and their likely effects... Furthermore, antitrust enforcers must be alert in these industries because the very nature of the “positive feedback” cycle means that monopolization may be accomplished swiftly. And, once achieved, the network effects that helped create dominance may make it more difficult for new entrants to dislodge the market leader than in other industries lacking network characteristics.”

The Justice Department’s Sherman Act case against Microsoft, which rested heavily on arguments involving network effects, was filed in May, 1998.

All this suggests that the Rohlfs paper reprinted here may have become one of those classics that is often cited but rarely read. I think that is unfortunate, but not because the network-effects literature than began in the mid-1980s is wrong in any technical sense or inapplicable to some markets. Nor do I think the policy concerns it raised are not relevant in those markets or that the enforcement stance expressed by Professor Shapiro in the thoughtful speech just cited is inappropriate in those markets.²⁰

The point is that the network-effects literature that began in the mid-1980s does not seem to have much to say about markets in which product or technology choices do not have long-lasting consequences—markets like those analyzed by Rohlfs. Without durability, expectations are not critical to decision-making, and lock-in is much less likely, particularly if multi-homing is possible.

And, as David Evans and I have argued,²¹ “almost every day” one now reads about markets of the sort analyzed by Rohlfs. Consider social networks, for instance. There are clearly direct network effects in these businesses: the value of being a participant in any particular social network depends on who else is participating. (There are also indirect network effects in these businesses, since participants attract advertisers.) But it is easy to switch between networks or to participate in multiple networks at the same time. Thus despite

THE POINT IS THAT THE NETWORK-EFFECTS LITERATURE THAT BEGAN IN THE MID-1980S DOES NOT SEEM TO HAVE MUCH TO SAY ABOUT MARKETS IN WHICH PRODUCT OR TECHNOLOGY CHOICES DO NOT HAVE LONG-LASTING CONSEQUENCES—MARKETS LIKE THOSE ANALYZED BY ROHLFS.

its early network-effect-enhanced advantages, MySpace has been almost totally eclipsed by Facebook, which used a university-based launch strategy that could have been cribbed from Rohlfs' discussion of launching a new service in a population consisting of multiple groups with strong intra-group affinities. Network effects, while present, were clearly not the only important factors in competition between MySpace and Facebook.

It is particularly interesting to note that video communications services, of the sort that Rohlfs could only hypothesize in the early 1970s, are now generally available on the Internet. And, as Rohlfs assumed, network effects plainly matter in choosing which service to use. But deciding to use one or another at any point in time has essentially no long-term implications. I currently use software for two such services on my computer and could no doubt access others if I saw benefits from doing so. I have no reason to care which service will prove more popular in the future, as I can use any one when it is worth using and ignore it otherwise.

ROHLFS, HOWEVER, CONSIDERED
A DIFFERENT SUBSET OF MARKETS
WITH NETWORK EFFECTS,
ONE THAT I BELIEVE IS BECOMING
MORE IMPORTANT IN PART
BECAUSE OF THE INTERNET.

As a final example, consider smart phone operating systems. As I write this, there is a vigorous struggle going on between Apple and Google, with Microsoft and RIM also engaged. But this is not the PC market in the 1990s: consumers buy new smart phones fairly often, and

the costs of switching between phones based on different operating systems do not seem to be significant. Thus, even though consumers don't generally multi-home in this market as they easily can for social networks and video calling services, product selection does not involve a very durable commitment, and competition does not look like a one-shot, winner-take-all affair.

Thus, while Jeffrey Rohlfs' paper has been influential in calling the attention of the economics profession to markets with network externalities, both economists and policy-makers have tended, until recently, to focus on a particular subset of those markets—those in which it is at least arguably the case that anticompetitive behavior during critical periods of winner-take-all competition may lead to undesirable and long-lasting outcomes. In such markets, especially close attention by antitrust authorities during those critical periods is appropriate. Rohlfs, however, considered a different subset of markets with network effects, one that I believe is becoming more important in part because of the internet. In these markets switching costs are not important, and the key decisions do not involve purchases of big-ticket, long-lived durables. Accordingly, participation decisions can easily be reversed. Moreover, multi-homing is often possible in these markets, so even at the individual level competition is not a winner-take-all matter even for short periods of time.

This is not to say that the markets considered by Rohlfs may not raise novel and interesting competition policy issues, since the presence of externalities and the potential for multiple equilibria imply the possibility of departures from the textbook norm. But it is far from clear that such markets deserve especially close antitrust scrutiny.²² In any case, I believe Rohlfs' analysis deserves to be read carefully and to be both extended and applied by economists and, further, that its implications should be carefully considered by antitrust enforcement agencies as they increasingly deal with markets for which that analysis is relevant. ▼

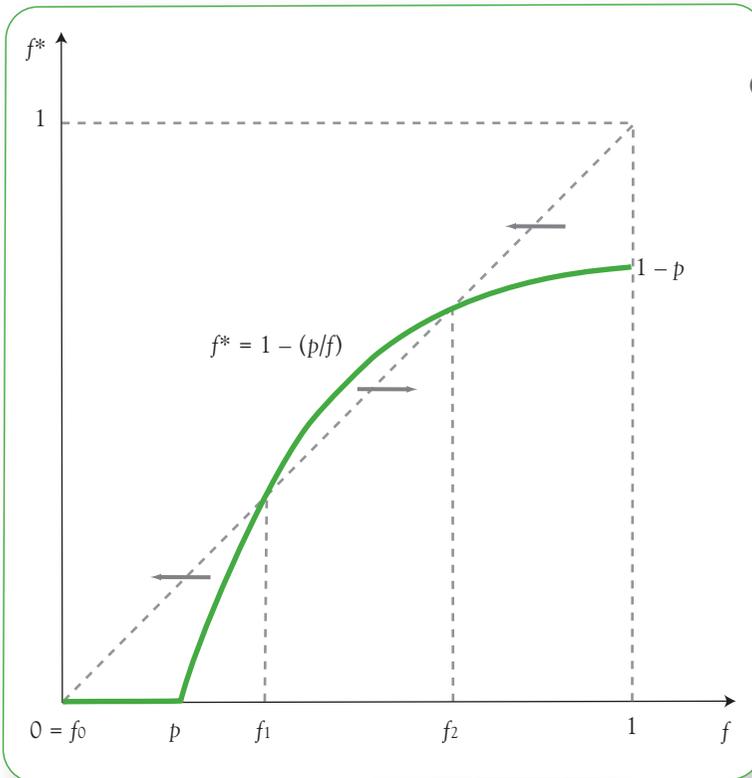


Figure 1

The Example from
Rohlfs' Section 3

- 1 Jeffrey Rohlfs, *A Theory of Interdependent Demand for a Telecommunications Service*, 5 BELL J. ECON. MGMT' T SCI. 5, 16-37 (1974), reprinted in this issue of COMPETITION POL'Y INT'L, (hereinafter "Rohlfs").
- 2 See, e.g., Jean-Charles Roche & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37 RAND J. ECON. 645-67 (2006).

- 3 Roland Artle & Christian Averous, *The Telephone System as a Public Good: Static and Dynamic Aspects*, 4 BELL J. ECON. MGMT SCI. 89-100 (1973) and Lyn Squire, *Some Aspects of Optimal Pricing for Telecommunications*, 4 BELL J. ECON. MGMT SCI. 515-525 (1973).
- 4 Rohlfs, *supra* note 1, at 24.
- 5 *Id.*
- 6 Suppose some equilibrium point, f_0 , is stable. Then df^*/df must be less than one at that point, so that $f^* > f$ for f just below f_0 , and $f^* < f$ for f just above f_0 . But then in order for f and f^* to be equal for some $f_1 > f_0$, f^* must rise faster than f near f_1 , since f^* is approaching f from below. Thus df^*/df must exceed one at f_1 , from which it follows that that equilibrium is unstable. Repeating the argument starting at an unstable equilibrium establishes the result.
- 7 Rohlfs, *supra* note 1, at 32.
- 8 *Id.* at 33.
- 9 E. Glen Weyl, *A Price Theory of Two-Sided Platforms*, 100 AMER. ECON. REV., 1642-72 (2010).
- 10 A common alternative in the internet world is to keep price to consumers at zero but to degrade quality by adding advertising.
- 11 If at some value of f , household i finds it optimal to subscribe, it must be that $w_i \geq p^*(f)/f$. This household will remain a subscriber as f increases if $p^*(f)/f$ does not increase, which is equivalent to the requirement that the elasticity of p^* with respect to f not exceed unity. Rohlfs requires $p^*(f)$ to be concave because he assumes that it passes through the origin.
- 12 All citation counts are as of March 10, 2011.
- 13 Todd Sandler & John T. Tschirhart, *The Economic Theory of Clubs: An Evaluative Survey*, 18 J. ECON. LIT. 1481-1521 (1980). This article has 330 Google Scholar citations.
- 14 Joseph Farrell & Garth Saloner, *Standardization, Compatibility, and Innovation*, 16 RAND J. ECON. 70-83 (1985) and Michael L. Katz & Carl Shapiro, *Technology Adoption in the Presence of Network Externalities*, 94 J. Pol. Econ., 822-841. These papers have 1662 and 1713 Google Scholar citations, respectively. Interestingly enough, the most frequently cited of the first-round network externality papers by these authors (Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 73 AMER. ECON. REV., 424-440 (1985)), which has 4113 Google Scholar citations, does not cite the Rohlfs paper. It cites instead a later paper that in turn cites Rohlfs' paper, along with the related earlier papers by Artle & Averous and by Squire, cited here in note 3, *supra*.
- 15 Michael L. Katz and Carl Shapiro, *Systems Competition and Network Effects*, Journal of Economic Perspectives 8, 93-115 (1994). This article has 1621 Google Scholar citations. Katz and Shapiro also note that the literature they survey considers compatibility decisions, but these are beyond the scope of the present discussion.
- 16 *Id.* at 93.
- 17 *Id.* at 105.
- 18 *Id.* at 113.

- 19 Speech delivered by Carl Shapiro before the American Law Institute and American Bar Association, January 25, 1996, p. 6; available at <http://www.justice.gov/atr/public/speeches/0593.pdf>.
- 20 But see Richard Schmalensee, *Antitrust Issues in Schumpeterian Industries*, 90 AMER. ECON. REV. PAPERS AND PROCEEDINGS, 471-5 (2000) and David S. Evans & Richard Schmalensee, *Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries*, INNOVATION POLICY AND THE ECONOMY, VOL. 6 (A. Jaffe, J. Lerner, & S. Stern, eds.) 2, 1-49 (2002).
- 21 David S. Evans & Richard Schmalensee, *Failure to Launch: Critical Mass in Platform Businesses*, 9 REV. NETWORK ECON., Article 1 (2010). That paper provides several brief case studies and generalizes the Rohlfs uniform calling analysis to platform businesses serving two customer groups, between which there are indirect network externalities. The nature of the critical mass constraint—"chicken and egg" or "chicken or egg"—is shown to depend on details of the adjustment process, and if there are multiple equilibria, stable equilibria and saddle-points alternate.
- 22 If network effects do lead to dominance, for instance, but that dominance can be quickly eroded by a better product either because participation decisions can be costlessly reversed or through multi-homing, the general presumption that network effects create troublesome entry barriers would not hold. Joe Farrell has pointed out (personal communication) that in some respects this sort of dominance resembles market-wide exclusive dealing contracts of very short duration.

A Theory of Interdependent Demand for a Communications Service

By Jeffrey Rohlfs*

The utility that a subscriber derives from a communications service increases as others join the system. This is a classic case of external economics in consumption and has fundamental importance for the economic analysis of the communications industry. This paper analyzes the economic theory of this kind of interdependent demand. We begin by defining “equilibrium user set” as a set of users consistent with all individuals’ (users and nonusers) maximizing their utilities. There are typically multiple equilibria at any given price and which equilibrium is attained depends partly on the static model, partly on the initial disequilibrium conditions, and partly on the disequilibrium adjustment process. Some general properties of equilibrium user sets are derived. Then we turn our attention to some specific models based on simple characterizations of communities of interest. The implications for pricing are discussed, with special reference to the problem of starting up a new communications service (e.g. a video communications service).

*This article was published in the Spring 1974 issue of the *Bell Journal of Economics and Management Science*. At the time, Jeffrey Rohlfs was a Member of the Technical Staff at Bell Labs, where his research interests included applied microeconomics, micro theory, and econometrics. He subsequently became Head of Economic Modeling Research at Bell Laboratories. Dr. Rohlfs received an A.B. in economics from Amherst College in 1965 and his Ph.D. in economics from M.I.T in 1969. In his introduction, Dr. Rohlfs wrote, “In the course of this study, I talked to many people, and their knowledge and ideas contributed to much of the analysis in this paper. M. Wish has been a collaborator in some previous related work, and he has greatly influenced my thinking. My analysis of the general problem was greatly stimulated by a discussion I had with E. Gilbert, who developed some preliminary results about the maximum equilibrium set. I have also had many profitable discussions with W. Ballamy, J. Berrier, A. Ciesielka, D. Deutsch, A. Gersho, E. Goldstein, A.H. McKeage, D. Mitra, and R. Sanders. F. Sinden, W. Taylor, N. Valcoff, M. Wilk, and E. Zajac provided helpful comments on previous written and oral presentations of the material.

I. Introduction

The utility that a subscriber derives from a communications service increases as others join the system. This is a classic case of external economics in consumption and has fundamental importance for the economic analysis of the communications industry. It suggests that although marginal cost pricing may be superior to allocated-cost formulae, it is still not completely appropriate.

This can be illustrated with respect to an historical policy of the industry: promoting universal service. This policy might be justified on the basis of marginal cost pricing, so long as new subscribers pay the incremental cost of expanding the system to accommodate them—even if they do not pay their “allocated” share of average costs. A still lower price, perhaps much lower, might be justified if the externalities are taken into account. The total benefits that *all* subscribers derive from the expansion of the service may be sufficient to justify the incremental costs—even if the subscribers are unwilling to pay the entire incremental costs.

Recently Artle and Averous¹ made what appears to be the first published analysis of these externalities in communications.² They formulate a simple model in which the incremental utility of the service to an individual depends only on the number of telephone subscribers—not on who they are. This is the uniform calling model discussed in Section 3 of this paper. They also assume that the cost of providing telephone service depends only on the number of subscribers. This enables them to derive and interpret the necessary conditions for a social welfare optimum. Their expression has some important similarities (but also some differences) for a social optimum with respect to a pure public good.

The authors then use these notions to develop a dynamic demand model. They show that interdependent demand can sustain continual growth in a stationary population with stationary income. The mechanism is as follows. New subscribers join. This increases the incremental utility of the service and induces marginal nonusers to join. That in turn induces further growth, etc., etc. The authors offer this as a possible explanation for the continual growth of telephone service observed in all empirical studies of the industry.

Squire studies the problem using a somewhat different model.³ He considers usage of the system as well as number of telephones and assumes that the cost of providing the service is a function of these two variables. Squire specifies individual demand curves (based on a fixed number of subscribers) for incoming and outgoing calls. This enables him to develop an expression for optimal usage of the system, based on a modified consumer-surplus concept. He then derives the optimal price per call (charged only to the person making the call) consistent with this optimal usage. He finally develops an expression for the optimal size of the system and the price per telephone consistent with this optimum.

This paper makes a much more detailed analysis of the demand side of the market than attempted by Artle and Averous or by Squire. We begin by defining

an “equilibrium user set” as a set of users consistent with all individuals’ (users and nonusers) maximizing their utilities. A basic result is that there are typically multiple equilibria at any given price. For example, a very small equilibrium user set may be consistent with utility maximization, since the smallness of the user set in itself makes the service relatively unattractive to potential users. However, a much larger user set may also be possible for the same population at the same price. In this case the largeness of the user set would make the service attractive and allow a high level of demand to be sustained. In any planning (public or private) for the communications service, special attention must be paid to which equilibrium user set is likely to be attained.

The next section of this paper develops a general theory of demand. It derives the following results:

- 1) The static model determines the attained equilibrium user set (at a given price) within a certain set of bounds.
- 2) A possibly narrower set of bounds (for a given price) is derived, given the initial user set.
- 3) Within the bounds defined in (2), the equilibrium attained (at a given price) depends entirely on the disequilibrium adjustment process.

The following two sections of the paper develop specialized models based on various simple characterizations of communities of interest. The simplest of all is the uniform calling pattern, which assumes that no one has any special community of interest (other than the entire population). This model is the only one in which the equilibrium theory can be developed in terms of the *number* of users, without paying attention to who they are. We can therefore define a demand curve, which turns out to have an inverted *U* shape. See Figure 1.

Zero demand is a stable equilibrium for all positive prices. The upward-sloping part of the inverted *U* consists of unstable equilibria and constitutes the “critical mass” of the service (at any given price). If the critical mass is exceeded demand expands to the downward sloping part of the inverted *U*. Points on the latter are stable equilibria and represent the maximum level of demand sustainable at a given price.

Unfortunately (for ease of analysis), the uniform calling pattern may not be very realistic. People typically belong to groups, each of which has a strong community of interest within itself. And they typically have a few principal contacts who alone account for a substantial part of their communication. These complications are briefly discussed in the section entitled “Nonuniform Calling Patterns.”

The final section of the paper discusses some implications of the preceding demand analysis for supply and pricing of the service. An important distinction is made between viability of service (existence of a nonnull equilibrium user set

that can be served with nonnegative profits) and the start-up problem (how to attain such a user set, starting from a small or null initial user set).

Viable nonnull equilibrium user sets (if they exist) are always superior to the null set from a static point of view. We can compare such sets to determine the static social optimum or the overall market equilibrium corresponding to a static supply model. However, this kind of analysis is incomplete and may be misleading without consideration of the start-up problem. Achieving the static optimal user set may require ruinous (albeit temporary) promotional costs.

Appropriate solutions to the start-up problem depend in large part on the demand model. In the uniform calling model, the start-up problem is simply a question of getting beyond the critical mass. Community of interest groups may make the practical start-up problem much easier, but they also introduce some special problems. If an individual's demand is contingent on a few principal contacts' being users, there may exist many small self-sufficient user sets. These allow the possibility of a long-term introductory program, in which the seller gradually expands the size and number of such sets.

This paper presents only a limited discussion of costs and supply. The reason is that costs of a communications service are very complex and merit a separate study in their own right. This is a very fruitful topic for future research.

II. General Theory of Demand

Let the population consist of n individuals. As in Artle and Averous' work, we define a set of binary variables:

$$q = \begin{cases} 0 & \text{if individual } i \text{ does not subscribe to the communications service} \\ 1 & \text{if the individual } i \text{ does subscribe to the communications service} \end{cases} \quad (1)$$

for $i = 1, \dots, n$.

We assume that there are also m other goods in the economy. To model interdependent demand, we specify a pair of utility functions for each individual:

$$U_i^0 = U_i^0(r_{i1}, \dots, r_{im}) \quad (2)$$

$$U_i^1 = U_i^1(q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n, r_{i1}, \dots, r_{im}) \quad (3)$$

where

U_i^0 = Utility of individual i if he does not subscribe to the communications service,

U_i^1 = Utility of individual i if he does subscribe to the communications service, and

r_{ij} = Consumption of (noncommunications) good j by individual i .

Equations (2) and (3) implicitly assume independent utilities with respect to all goods in the economy *other* than the communications service in question. In addition we make the usual monotonicity assumptions:

$$\frac{\partial U_i^k}{\partial r_{ij}} \geq 0 \text{ for all } j \text{ and } > 0 \text{ for some } j; \text{ and} \tag{4}$$

$$U_i^0 \leq U_i^1 \tag{5}$$

for all $i, k, q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n, r_{i1}, \dots, r_{im}$.

We also make a specialized assumption applicable to a communications service:

$$\frac{\partial U_i^1}{\partial q_m} \tag{6}$$

for all $i \neq w, q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n, r_{i1}, \dots, r_{im}$. That is, a subscriber's utility never decreases as additional individuals subscribe (and none drop out).

This seems like a reasonable working assumption. We can, of course, imagine some exceptions: e.g., the value of the service to others would probably be lessened if a large number of life insurance salesmen subscribed to the service to solicit other subscribers. However, we assume that such occurrences are the exception rather than the rule—that, in general, the availability of a communications link is not detrimental to either party.

We assume utility maximization, which we analyze in two steps. (1) We evaluate the maxima of U_i^0 and U_i^1 (with respect to r_{i1}, \dots, r_{im}) subject to individual i 's budget constraint. Let us denote these maxima as \hat{U}_i^0 and \hat{U}_i^1 . (2) We then compare \hat{U}_i^0 and \hat{U}_i^1 to see if the individual demands the communication service. This defines a demand variable for each individual:

$$q_i^D = \begin{cases} 0 & \text{if } \hat{U}_i^0 > \hat{U}_i^1 \\ 1 & \text{if } U_i^0 \leq \hat{U}_i^1 \end{cases} \tag{7}$$

for $i = 1, \dots, n$.

The basic methodology of this paper is to ignore interrelationships between the communications market and other markets and concentrate on relationships *within* the communications market. Thus, we make the *ceteris paribus* assumption that prices of all goods other than the communications service are fixed and that each individual has a fixed budget constraint. This allows us to express the demand variables as functions of price and the set of subscribers:

$$q_i^D = q_i^D(p, q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n) \tag{8}$$

for $i = 1, \dots, n$, where p = the price of the communications service.

It follows from previous assumptions that all the q_i^D are monotonically decreasing (equality allowed) with respect to p . That is, an increase in p can never change q_i^D from 0 to 1; a decrease in p can never change q_i^D from 1 to 0. However a change in p may have no effect on q_i^D . It also follows from previous assumptions that all the q_i^D are monotonically increasing (equality allowed) with respect to all q_w ($w \neq i$).

A. EQUILIBRIUM USER SETS

Naturally, there is a correspondence between demanding the service and being a subscriber. We define an *equilibrium user set* as a set of users such that

$$q_1 = q_1^D(p, q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n) \quad (9)$$

for all i .⁴ Thus, in equilibrium all users demand the service; all nonusers do not demand it.

Equation (9) defines equilibrium with respect only to the demand side of the market. It describes user sets that are consistent with utility maximization at a given price. These constitute necessary but not sufficient conditions for an overall market equilibrium. The latter additionally requires that the user set and price be consistent with some specified model of supply behavior.

For fixed $p = \bar{p}$, equations (9) are a system of n equations in n binary variables. Such a system does *not* generally have a unique solution. In fact, unique solutions did not arise in any of the simple models investigated in this paper (except in the trivial case where price is so high that there can never be any demand at all).

Consequently, the equation

$$q = q^D, \quad (10)$$

where

$$q = \sum_{i=1}^n q_i \text{ and}$$

$$q = \sum_{i=1}^n q_i^D$$

may be indeterminate (for fixed \bar{p}). That is, it may either hold or fail to hold depending on which set of users constitutes the sum q .

For this reason, the general theory of interdependent demand cannot be developed in terms of the sum q . It is necessary to work with the individual q_i . The basic analytical concept is not the demand curve—i.e. the equilibrium pairs (q, p) —but rather equilibrium user sets.

B. DISEQUILIBRIUM ANALYSIS

Given that several equilibrium user sets exist for a given price (*ceteris paribus*), it is important to know which ones (if any) are most likely to occur. This requires analyzing what happens if the market is initially in disequilibrium. Our procedure is as follows. We specify a very general disequilibrium adjustment process. We then investigate the extent to which the user sets resulting from this process depend on the static model, the extent to which they depend on the initial disequilibrium conditions, and the extent to which they are indeterminate, depending on a more detailed specification of the adjustment process.

In this section, we restrict our attention to the demand side of the market and assume a given price for the communications service. We further assume that adjustments of consumption in other markets can be made rapidly and costlessly. This seems like a reasonable simplifying assumption, allowing us to analyze disequilibria in the communications market without considering possible disequilibria in the rest of the economy.

Now suppose there is an arbitrary user set. It may be based on utility maximization for current or previous states of the world, past selling efforts of the supplier of the service, or anything else. We assume that adjustments to this user set occur according to the following adjustment process. (1) An individual in equilibrium ($q_i^D = q_i$) never changes his status from user to nonuser or vice versa. This is reasonable, since such a change would always reduce his utility (except in the knife-edge case where $\hat{U}_i^0 = \hat{U}_i^1$, in which case the change in status has no effect on utility). (2) The length of time an individual can remain continually in disequilibrium ($q_i^D \neq q_i$) is bounded. He eventually must change his status. This is also reasonable, since the change always increases his utility (except in the knife-edge case where $\hat{U}_i^0 = \hat{U}_i^1$, in which case the change in status has no effect on utility).

The adjustment process is essentially a model of utility maximization with inertia. It is very general in that it makes no assumption about the speed of adjustment. This speed may vary from individual to individual. It may depend on the user set or actions of the seller. Or it may change over time.

A limitation of this process is that it does not allow individuals to collude and subscribe together. This is relatively unimportant if an individual's demand is contingent on a large user set, since such collusion would be difficult with very large groups. However (as will be seen later) we do have to consider relaxing the assumption in models where an individual's demand is contingent on a few of his principal contacts' being users.

It is important to note that the adjustment process does *not* necessarily converge to an equilibrium user set. Consider the following example. A demands the service if and only if B is a user; B demands the service if and only if C is a user; C demands the service if and only if A is a user. Suppose the initial user set is A.

One possible version of the adjustment process is as follows. C joins because A is a user. Then A disconnects because B is not a user. Then B joins because C is a user. Then C disconnects because A is not a user. Then A joins because B is a user. Then B disconnects because C is not a user. We are now back to the original user set, and the process can be repeated indefinitely.

Nevertheless, the user sets resulting from the adjustment process can be bounded, as shown in the following theorems:

Theorem 1: If the initial user set is the entire population, the adjustment process can only *remove* individuals from the user set; no individual can ever be added who has previously dropped out.

Proof: If the entire population is an equilibrium user set no one is added or removed, and the theorem is satisfied. If the entire population is *not* an equilibrium set, let r_1, r_2, \dots represent the sequence of individuals who change status. (If individuals change status simultaneously, we list them in arbitrary order.) Now r_1 must be a removal (not an addition), since the entire population consists of users, and there is no one left to be added. Given that r_1, \dots, r_k are removals, r_{k+1} must also be a removal for the following reason. The only possible additions would be the individuals r_1, \dots, r_k . But all of these dropped out (and therefore did not demand the service) when the user set contained the current user set. Thus, they cannot demand the service according to the monotonicity assumption.

It follows that all of the r_i must be removals. Q.E.D.

Theorem 2: If the initial user set is the population, the adjustment process converges to an equilibrium user set in finite time.

Proof: The process must converge in finite time for the following reason. All changes of status are removals. Since only n (the size of the population) individuals can be removed, there are at most n changes in status. These must all occur in finite time.

After all the changes in status occur, no user can fail to demand the service, for otherwise the process would continue. Moreover, no nonuser can demand the service because of the monotonicity condition. Thus, the final user set is an equilibrium user set. Q.E.D.

Theorem 3: If the initial user set is the entire population, the adjustment process converges to the union of all equilibrium user sets regardless of the order of removals.

Proof: Let X be an arbitrary equilibrium user set; let R be the equilibrium result after individuals r_1, \dots, r_k have been removed according to the adjustment process. X cannot contain r_1 , since r_1 was removed when the user set was the entire population (and hence contained X). Since X does not contain r_1 , it cannot contain

r_2 either. (The entire population minus r_1 contains $X - r_1$.) Nor can X contain r_3, \dots, r_k . Thus, $X \subseteq R$.

But X is an arbitrary equilibrium user set. Thus, R contains all equilibrium user sets. Since R is itself an equilibrium user set, it is the union of all equilibrium user sets. Q.E.D.

This set will hereafter be referred to as the “maximum equilibrium user set.”

By entirely symmetrical reasoning we can establish the following. If the initial user set is null,

1. the adjustment process can only add individuals, and no one is ever removed who previously joined;
2. the adjustment process converges to an equilibrium user set in finite time;
3. the adjustment process converges to the intersection of all equilibrium user sets regardless of the order in which individuals are added.

This set will hereafter be referred to as the “minimum equilibrium user set.”

It is important to note that the minimum and maximum equilibrium user sets need not be the same. For example, consider the following simple model. Each individual demands the service if three of his five principal contacts are users. The minimum equilibrium user set is null. There are no users; so no one has three of his five principal contacts as users; so no one demands the service. The maximum equilibrium set is the entire population. Everyone has all five of his principal contacts as users; so everyone demands the service. In addition, there may be any number of equilibrium user sets between these two extremes, depending on the distribution of principal contacts.

In this example, the static model tells us practically nothing about what equilibrium will actually be attained. We only know that it will be zero or 100 percent or somewhere in between. This is an extreme case but in general the static model determines the actual equilibrium only within certain bounds—the minimum and maximum equilibrium sets. Moreover, in all of the models investigated in this paper there exists the possibility that these bounds may be far apart. In a practical situation, this difference may mean the difference between marketing success and failure.

The above theorems show that if the initial user set is sufficiently large, convergence to the maximum equilibrium user set is assured (according to the assumed adjustment process). If the initial user set is sufficiently small, convergence to the minimum equilibrium set is assured. For intermediate initial disequilibrium user sets, the actual equilibrium attained may also depend on a more detailed specification of the adjustment process than given above.⁵ It may be crit-

ical whether or not the disequilibrium nonusers subscribe before the disequilibrium users drop out.

C. PARTICULAR INITIAL USER SET.

The minimum and maximum equilibrium user sets provide bounds on user sets that are possible for *any* version of the adjustment process described above.⁶ This subsection provides bounds on user sets that can be attained from a *particular* initial user set for *any* version of the adjustment process.

Let S be an arbitrary initial user set. We now define the following two adjustment sequences.

Optimistic sequence

1. First all nonusers who demand the service subscribe in arbitrary order but no users drop out. This converges to the same user set \bar{S} , irrespective of the order in which individuals subscribe. (Proof is analogous to that of Theorem 3.)
2. Then all users who do not demand the service drop out in arbitrary order. This converges to the same user set S , irrespective of the order in which individuals drop out. Moreover, \bar{S} is an equilibrium user set. (Proof analogous to Theorems 2 and 3.)

Pessimistic sequence

1. First all users who do not demand the service drop out in arbitrary order, but no nonusers subscribe. This converges to the same user set \underline{S} , irrespective of the order in which individuals drop out. (Proof analogous to Theorem 3).
2. Then all nonusers who demand the service subscribe (in arbitrary order). This converges to the same user set \underline{S} , irrespective of the order in which individuals subscribe. Moreover, \underline{S} is an equilibrium user set. (Proof analogous to Theorems 2 and 3.)

Now let R_1, R_2, \dots be a sequence of user sets resulting from applying an arbitrary version of the adjustment process to S . As previously discussed, this sequence need not converge to an equilibrium user set. However, we can place the following bounds on the sequence:

1. $\underline{S} \subseteq R_i \subseteq \bar{S}$ for all i . This follows directly from the monotonicity assumption.
2. After some finite period of time, $\underline{S} \subseteq R_i \subseteq \bar{S}$ for all i .

Proof: Let x_1, \dots, x_k be a sequence of individuals who drop out in part (2) of the optimistic sequence. Since x_i does not demand the service given user set \bar{S} , he cannot demand it given any user set R_i . Thus, if x_i is a user, he is continually in

disequilibrium. He must drop out in finite time and can never thereafter rejoin. Once x_i drops out we can apply the same reasoning sequentially to x_2, \dots, x_m . Thus, after some finite period of time, all $R_i \subseteq \bar{S}$. The proof that $S \subseteq \bar{R}_i$ is exactly symmetrical. Q.E.D.

Thus, the optimistic and pessimistic sequences define bounds on user sets attainable from a particular initial user set. These bounds may (or may not) be considerably narrower than the bounds provided by the minimum and maximum equilibrium sets. In any event, within these bounds the equilibrium use set attained depends entirely on a detailed specification of the adjustment process.

D. ADDITIVE UTILITIES.

In order to proceed further we must make more assumptions. To simplify the problem we propose a model of additive utilities. That is, we define a vector \underline{f} and a matrix \underline{V} such that

$$U_i^0 = f_i(r_{i1}, \dots, r_{im}) \tag{11}$$

$$U_i^1 = f_i(r_{i1}, \dots, r_{im}) + \sum_{j \neq i} v_{ij} q_j, \tag{12}$$

where v_{ij} ($i \neq j$) is the incremental utility to individual i of a communications link with individual j , ($v_{ij} \geq 0$).

The additive model assumes that these incremental utilities do not depend on consumption of other goods or on other communications links available to the individual. These do seem to be reasonable simplifying assumptions, but there are some problems with them. The growth of telephone service has had fundamental effects on social and business customs, and these would not be captured in an additive model. It has also resulted in substantial changes in communities of interest, which are assumed to be fixed in equation (12). However, the additive model would be commensurately better for analyzing smaller differences in market penetration or for analyzing a service that does not provide so revolutionary an improvement in communications as did the invention of the telephone.

Equation (12) also assumes that the service has no value except to communicate with others who have the service. The service is worthless if no one else subscribes. This assures that the null set is an equilibrium user set at any positive price.

This assumption sounds reasonable enough, but there are some possible exceptions. An individual may have noncommunications applications for the hardware. If the service is new, he may find it prestigious or derive self-satisfaction from being an innovator. However, these kinds of considerations go beyond the scope of this paper.

The additivity assumption is quite useful and allows us to derive a convenient expression for q_1^D as shown below.

The maxima \hat{U}_i^0 are defined by the *ceteris paribus* conditions and do not depend on anything in the communications industry. Maximizing equation (12) with respect to r_{i1}, \dots, r_{im} , subject to individual i 's budget constraint, we obtain

$$\hat{U}_i^1 = \hat{U}_i^0 - h_i(p) + \sum_{j \neq i} v_{ij} q_j \tag{13}$$

for some function h_1 where $h_1(0) = 0, h_1(p) > 0$ for all i .

It follows that

$$q_i^D = \{0 \text{ if } \sum_{j \neq i} v_{ij} q_j < h_i(p) \\ 1 \text{ if } \sum_{j \neq i} v_{ij} q_j \geq h_i(p)\} \tag{14}$$

where $v_{ij} \geq 0$ for all i, j .

We also assume constant marginal utility of money for a given individual. This means that $h_i(p)$ is a linear function:

$$h_i(p) = b_i p. \tag{15}$$

We can therefore write equation (14) as follows:

$$q_i^D = \{0 \text{ if } \sum_{j \neq i} w_{ij} q_j < p \\ 1 \text{ if } \sum_{j \neq i} w_{ij} q_j \geq p\} \tag{16}$$

where $w_{ij} = \frac{v_{ij}}{b_i}$ for all $i \neq j$.

E. FURTHER SIMPLIFICATION.

Both the monotonicity and the additivity assumptions greatly simplify the problem (at some cost in realism). However, we still must deal with the matrix \underline{V} which is the size of the population squared. Thus for a city with a population of one million, \underline{V} would have one trillion entries. Clearly, further simplification is required. The following two sections consider some possibilities for breaking the problem down to manageable size.

III. Uniform Calling Pattern

The preceding section began by considering the problem in its full generality and considered some reasonable kinds of simplifying assumptions. We now take the opposite approach, beginning with a very simple model and then relaxing assumptions to make the model more complicated and realistic.

In this section we assume that all the (off-diagonal) elements in any single row of \underline{V} are equal. This implies that no one has any special community of interest

other than the entire population. The *number* of subscribers affects an individual's demand, but he does not care who these subscribers are.

This may in fact be a reasonable approximation (for some purposes). We might reason that the individual communicates with a large number of people during the course of a year, many of whom he does not know in advance. The number of users may be as good a proxy as any for the incremental utility he derives from the service.

However, it is also true that most people belong to groups, each of which has a community of interest within itself. They also typically have a few principal contacts with whom they communicate more than with others. Thus, their demand for a communications service would depend on how many members of their community of interest group and which of their principal contacts subscribe to that service.

In any event, the uniform calling model seems like a good place to begin developing the theory. (This model is also adopted by Artle and Averous and Squire.)⁷ It allows some strong results to be derived and provides some useful insights about interdependent demand. Results from the uniform calling model also provide convenient reference points for analyzing more complex models, which are briefly discussed in the next section of this paper.

The uniform calling model allows us to write equation (15) as follows for a large population:

$$q_i^D = \begin{cases} 0 & \text{if } fw_i < p \\ 1 & \text{if } fw_i \geq p \end{cases}, \quad (17)$$

where f = the user fraction (q/n), and

$$w_i = \sum_{j \neq i} w_j.$$

This in turn allows individuals to be ordered in terms of their demand for the service. That is, if $w_i \geq w_j$ (i 's demand exceeds j 's), individual i is in every equilibrium set that contains j .

A. DEMAND CURVE.

Since individuals can be ordered as above, every equilibrium user set consists of *all* individuals (i) for whom $w_i \geq$ some K . Similarly, for any q , there is at most one equilibrium user set with q members; i.e., the q people with the highest value of w_i . (If more than one person has the minimum w_i in the user set, all persons with that w_i must be in the user set for it to be an equilibrium.)

Thus, every equilibrium user set can be uniquely characterized by q , the number of members in it. We can develop the equilibrium theory for this model in terms

of the sum q , without specifying the individual elements q_i . In particular, we can define a demand curve; i.e., the locus of all the pairs (q,p) for which there exists an equilibrium user set (which would have to be the q people with highest w_i).

This gives us a convenient way of looking at the relationship between price and the equilibrium user sets. However, it is important to note that equation (10) is still indeterminate, and we must be careful in applying the demand curve in disequilibrium situations.

B. AN EXAMPLE.

Before discussing the general properties of such a demand curve, let us consider a specific example. Suppose the population is large, and w_i is distributed uniformly between 0 and 100 over the population. For the marginal individual

$$w_i = 100(1 - f). \tag{18}$$

For an equilibrium at $0 < f < 1$, fw_i for the marginal individual must equal p . [See equation (17).] Thus, the demand curve is the locus of points where

$$100f(1 - f) = p \tag{19}$$

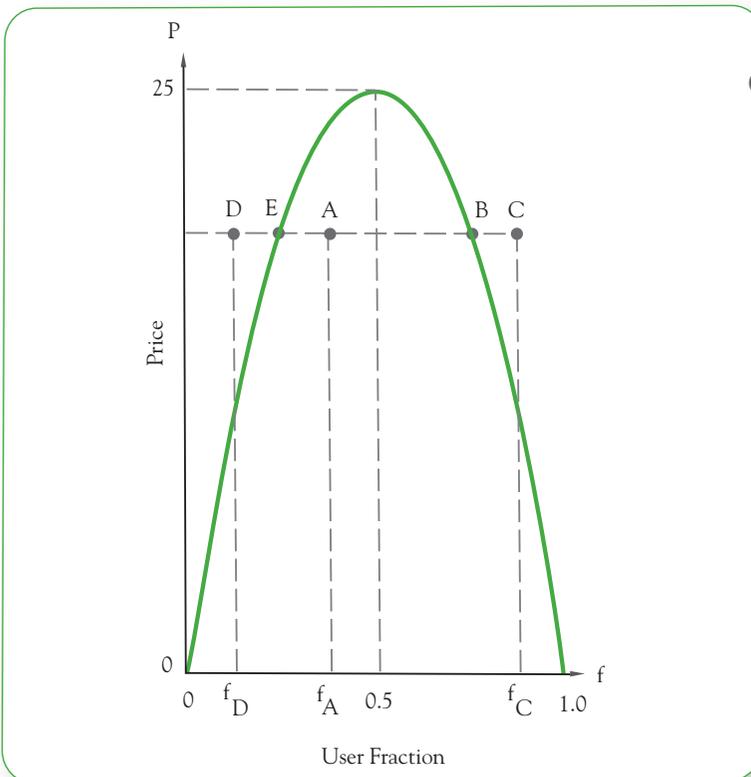


Figure 1

Demand Curve
for Uniform
Calling Pattern

As previously mentioned⁸ the null set ($f = 0$) is an equilibrium user set for all $p > 0$. For an equilibrium at ($f = 1$), p must be less than or equal to fw_i for all individuals. But the minimum of w_i and hence fw_i is 0. Thus, the only equilibrium is $p = 0$.

Figure 1 shows the demand curve. It consists of the entire positive p -axis plus an inverted parabola going through $(0,0)$ and $(1,0)$ and having a maximum at $(0.50,25)$.

The maximum equilibrium set is the right-hand side of the parabola for $0 < p \leq 25$; it is null for $p > 25$.

For small p , there is an enormous difference between the minimum and maximum equilibrium user sets. Thus, the actual equilibrium attained (for small p) depends critically on the initial disequilibrium condition and the disequilibrium adjustment process.

C. DISEQUILIBRIUM ANALYSIS.

The following analysis of disequilibrium is based on the adjustment process proposed above.⁹ As before, we assume a fixed price and restrict our attention to the demand side of the market. We first consider the special case in which the initial disequilibrium users are those with the highest w_i . We then consider the general case by examining arbitrary perturbations from an initial equilibrium.

D. INITIAL USERS HAVE HIGHEST w_i .

This subsection assumes that the initial user set consists of all individuals for whom $w_i \geq$ some K . This is necessarily true if the users form an equilibrium user set for any \bar{p} . Thus, the results apply to any disequilibrium brought about by a price change from an initial equilibrium.

Suppose we are originally in disequilibrium at A (in Figure 1), underneath the parabola. Given the user fraction f_A , the equilibrium price is higher than the actual price. All users are satisfied, but some nonusers would prefer to become users. If p remains constant, the user fractions will ultimately increase to B .

Suppose we are originally in disequilibrium at C . Given f_C , actual price exceeds equilibrium price and f declines to B .

Suppose we are originally in disequilibrium at D . Given f_D , actual price exceeds the equilibrium price. So, f declines. As f declines, the discrepancy between actual and equilibrium price increases until the market achieves equilibrium at $f = 0$.

In all these cases, the order in which individuals join or drop out is immaterial. The optimistic and pessimistic sequences are equivalent, and all versions of the adjustment process converge to the same equilibrium user set.

In general, the positive p -axis and the downward sloping part of the parabola in figure 1 consist of stable equilibria. The upward sloping part of the parabola consist of unstable equilibria.

The upward sloping part of the parabola can be regarded as a “critical mass” for the service. That is, for any positive price below the maximum of the parabola, the market must be forced to some initial disequilibrium beyond the critical mass before the service can grow by itself. The higher the price, the higher is the critical mass.

E. ARBITRARY PERTURBATIONS FROM EQUILIBRIUM.

The preceding analysis does *not* necessarily apply for arbitrary initial conditions. An initial user set of γ people may converge to very different equilibria depending on who those γ people are and on a more detailed specification of the adjustment process (than given above¹⁰).

For example, suppose the initial user set of γ people contains none of the γ people with highest w_i . Suppose it contains the people with the next γ highest values of w_i . Even though the initial market penetration is γ , the market can achieve a critical mass of 2γ if the γ nonusers with highest w_i all subscribe before any of the initial users drop out (optimistic sequence). However, if all the initial users who do not demand the service drop out before any nonusers subscribe (pessimistic sequence), the market may fail to achieve a critical mass much lower than γ .

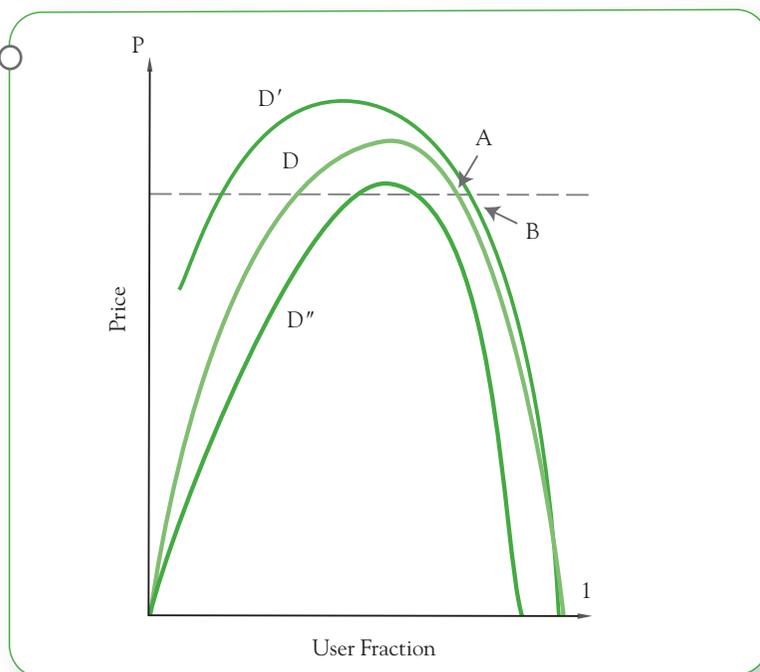
The results of the preceding subsection do apply for small but arbitrary perturbations (in the user set) from an original stable equilibrium. For example, suppose the market is originally in the stable equilibrium at A in Figure 2, and a set (R) of nonusers subscribes. We can analyze this by constructing a demand curve conditional on all members of R being users (D' in Figure 2). The perturbation of R becoming users may cause additional users to subscribe. However, no matter what the adjustment process is (subject to the rules laid down above¹¹), demand can never expand beyond B . And at B (or any point between B and A), only people in the *original* equilibrium set demand the service. Thus, the market eventually goes back to A (so long as the price remains fixed).

In the general case the perturbation may involve users dropping out as well as nonusers joining. We analyze this by constructing D' as before and D'' , demand conditional on those who drop out being nonusers. The market might converge to $f = 0$, if the perturbation is large and brings demand below critical mass. However, if the perturbation is sufficiently small, D'' will be sufficiently close to D that demand cannot go below critical mass. Thus, the market must return to the initial equilibrium.

This reasoning also applies to the stable equilibria at $f = 0$. If the perturbation is sufficiently small, D' will be sufficiently close to D that critical mass cannot be achieved, and the market must return to the initial equilibrium.

Figure 2

Arbitrary
Perturbations
from Equilibrium



F. GENERAL PROPERTIES OF THE MODEL.

Some properties of the above example apply generally to all uniform calling models. The entire positive p -axis always consists of stable equilibria. The demand curve always has an upward-sloping part, which constitutes the critical mass for the service for initial users sets with maximal w . It always has a downward sloping part (perhaps vertical), which consists of stable equilibria. However, both parts need not be unique, and the demand curve may be jagged. This allows the possibility of many stable equilibria for a given price.

IV. Nonuniform Calling Patterns

This section considers some models that are more complex than the uniform calling model. Some specific results are presented, but they are naturally not so strong as those of the previous section. Our primary objective is to point out the analytical complexities in such models and suggest some ways of dealing with them.

A. COMMUNITY OF INTEREST GROUPS.

Suppose the population consists of k groups ($k \leq n$). We assume that an individual has the same community of interest with everyone in the same group. However, this community of interest may be different for different groups. Mathematically, we assume that if individuals j and m are in the same group, $v_{ij} = v_{im}$ for all i .

B. DISJOINT GROUPS.

The simplest case is disjoint groups. That is, $v_{ij} = 0$ unless i and j are in the same group. In such a model, we can consider each group as a separate population, and all the analysis of the previous section carries over. A critical mass can be defined for each group in terms of market penetration within that group. For given \bar{p} the maximum possible number of stable equilibria is 2^k (unless the demand curve for some groups is jagged). The equilibria are characterized by which of the k groups achieve their critical mass.

C. JOINT GROUPS.

We now consider the case of joint groups, where v_{ij} does not necessarily equal zero for i and j in different groups. In this model, the incremental utility of the service to an individual is a function of g_i (and p). This is illustrated in Figure 3 for the case of the two groups.

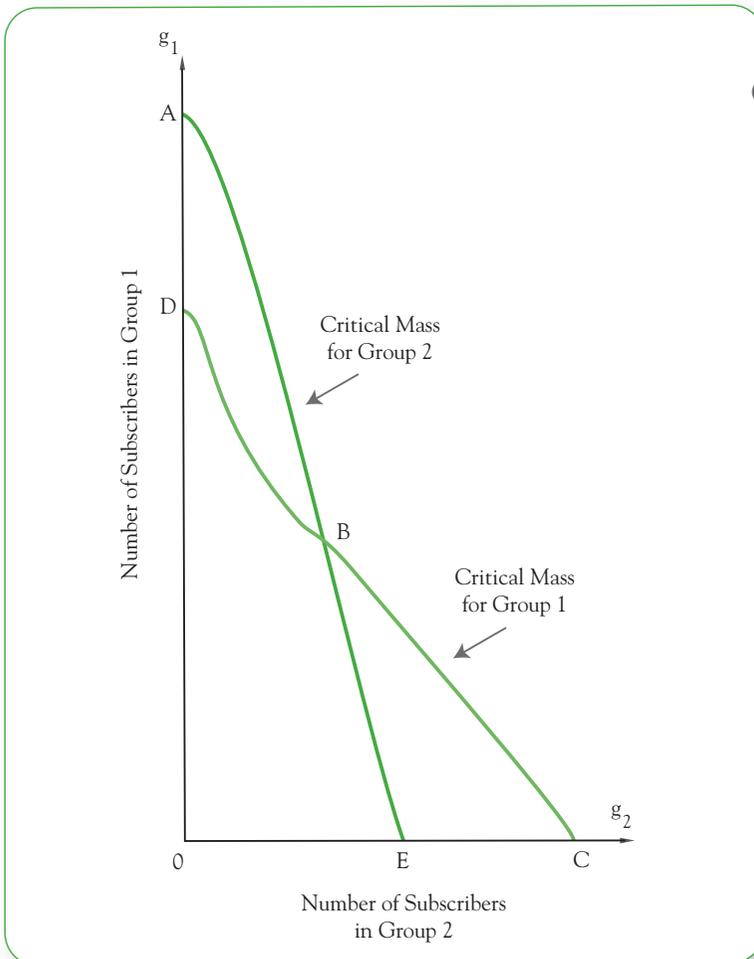


Figure 3

Critical Masses for Model with Communities of Interest

If the initial user set in Figure 3 consists of individuals with the highest w_i in each group (a weaker condition than requiring initial users to have the highest w_i in the population), we have the following result. If all groups achieve their critical mass (initial market penetration outside *ABCO* in Figure 3), the service will expand to the maximum equilibrium set. If no group achieves its critical mass (initial market penetration within *DBEO* in Figure 3), the service will collapse to the minimum equilibrium set. If some groups achieve their critical mass but others do not (initial market penetration within *ABD* or *BCE* in Figure 3), the service may expand to the maximum equilibrium set, collapse to the minimum equilibrium set, or achieve equilibrium somewhere in between. Which of these occurs depends on the parameters of the static model, the initial market penetration in each group, and the disequilibrium adjustment process. An upper bound on the number of stable equilibria is 2^k (unless the demand surface is jagged). Each equilibrium is characterized by which of the k groups achieve their critical mass and which do not.

D. FURTHER REFINEMENTS.

Introducing further refinements into the model is straightforward. As before, we somehow divide the population in k groups such that within each group individuals can be ordered in terms of their demand for the service. For greater realism, we could have a large value for k , but that has the drawback of requiring us to deal with a k -dimensional quantity vector and to contend with the possibility of many different equilibria (for each price).

E. FEW PRINCIPAL CONTACTS.

An individual's demand may depend primarily on which of his few principal contacts are users. A basic analytical tool for studying such demand is the "self-sufficient" user set; i.e., a set of individuals, each of whom demands the service conditional on all others in the set being users. An equilibrium user set must, of course, be self sufficient, but the converse is not necessarily true. Someone outside the self-sufficient set may demand the service if everyone in the set has it.

All self-sufficient sets necessarily belong to the maximum equilibrium user set. Moreover, if the entire self-sufficient set is contained in the initial disequilibrium user set, then the entire self-sufficient set is necessarily part of the final equilibrium user set.

In any practical problem, we could never hope to have a complete empirical list of principal contacts. The way to proceed in such cases is to specify a probability distribution which indicates (approximately) how likely various configurations of principal contacts are. This leads to some interesting combinatorial analysis but goes beyond the scope of this paper.

V. Some Implications for Supply and Pricing

Costs of providing a communications service depend on the consistency of the user set as well as its size. These costs are in themselves fully as complex as interdependent demand. Such complexities go beyond the scope of this paper, which is a study of interdependent *demand*. Nevertheless, it is useful to look at some implications of the preceding sections for supply and pricing of the service. We assume that the service is provided by a (regulated) monopoly and specify a very general cost function:

$$C = C(q_1, \dots, q_n), \quad (20)$$

where C is strictly monotonic (equality *not* allowed) in all its arguments.

We investigate various kinds of pricing strategies, some involving short-run losses. However, we assume that in (long-run) equilibrium, the monopoly must earn nonnegative profits. Some of the pricing strategies considered involve discriminatory pricing, but we do not assume that perfect discrimination is necessarily possible. Indeed, perfect discrimination would surely not be possible in any realistic situation. Consequently, the nonnegative profit restriction may be inconsistent with Pareto optimality. It is nevertheless consistent with existing real-world institutions.

We can now make a crucial distinction for planning supply of a communications service; i.e., the difference between “viability of the service” and “the start-up problem.” Viability is determined solely by the static model. It means that there exists a nonnull equilibrium user set that can be served with nonnegative profits. (We also refer to such a user set as “viable.”) The start-up problem is a dynamic consideration. It refers to the costs and practical difficulties of attaining a viable user set, starting from a small or null initial user set.

From a static point of view, any viable user set is superior to the null user set (in the sense that the supplier of the service and all users are at least as well off as before and possibly better off. Nonusers’ utilities are unchanged).¹² If there are several viable user sets, they can be compared to determine the social optimum (subject to the nonnegative profit restrictions). We can also determine the overall market equilibria consistent with various static supply models.

However, this kind of static analysis is incomplete and may in fact be misleading. We must also consider the dynamic aspects; i.e. the start-up problem. If the initial user set is small or null, the static social optimum may require ruinous (albeit temporary) promotional costs. Thus, all things considered, a smaller user set, or perhaps even the null set, may be superior.

The remainder of this section discusses various possible solutions to the start-up problem. We consider the case of a new service (e.g., a video communications service) and assume that the initial user set is null. This is necessarily an equilib-

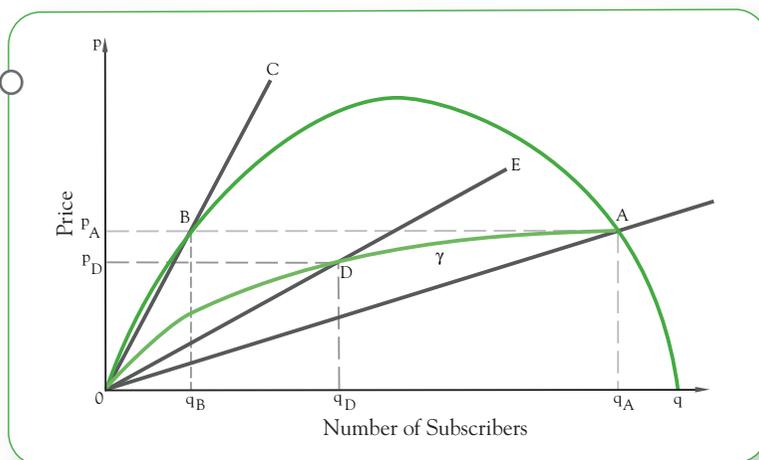
rium user set at all positive prices. Thus, the service, even though viable, cannot get started by itself. It requires some positive action by the seller, probably involving temporary losses. The next three subsections consider some possible start-up strategies in the context of the various demand models we have studied in the previous sections.

Although we assume in this analysis that the product is viable, it is worth noting that in real life the seller would have no such guarantee. There is always a risk involved in introducing a new product or service. The seller generally faces certain losses when the product or service is first introduced and the *prospect* of future profits. In a regulatory environment, the supplier must consider how the regulators will respond if the service succeeds and how they will respond if it fails.

A. UNIFORM CALLING PATTERN.

Let us assume that the demand curve is a (nonjagged) inverted U, and suppose the desired nonnull equilibrium set is A in Figure 4. The long-run optimal price is P_A , but some method must be contrived to get beyond the critical mass.

Figure 4
Solving the Start-Up Problem



B. DIRECT APPROACH.

The most direct approach is to give the service free to a selected group of people for a limited time. For this method to succeed, the initial user set must, of course, be sufficiently large to achieve critical mass. Half measures are worse than useless. If critical mass is not achieved, the whole effort will be a complete failure, and demand will eventually contract to zero.

The success of this approach may also depend on how the initial user set is selected. The discussion above¹³ shows that the optimal initial user does not necessarily consist of those with maximal w_i . (In the example, an alternative choice

of y people allowed a critical mass of $2y$ to be achieved.) However, if any other set is chosen, there should be some assurance that the initial users will not discontinue the service before those with higher w_i subscribe. If they do discontinue, the start-up effort will have failed to achieve anything.

C. LOW INTRODUCTORY PRICE.

Another way to start up the service is to have a low introductory price. The price could then be raised as the number of subscribers increased. There are any number of ways this could be accomplished. We can represent the introductory program as an expansion path in (q,p) space; e.g., γ in Figure 4. The expansion path shows how p increases as q increases.

The expansion path must pass through the origin to get the service started in this simple model. If, in addition, the expansion path is always concave downward, the introductory program has a very desirable property. Regardless of the order in which individuals enter the market, no individual ever subscribes and then later discontinues service after price and quantity rise along the expansion path. The proof is as follows. Suppose an individual enters at D . That means he is willing to pay p_D to have the service available for q_D of his communication. It follows from the uniform calling assumption that he would be an equilibrium user at any point along the straight line OE through D . But the concavity assumption assures that the expansion path γ is always below OE for $p > p_D$, $q > q_D$. Thus, if the individual joins at D , he cannot drop out at any point on γ to the right of D . Q.E.D.

The above condition can be very important if the cost of connecting an individual to the network is large.

Perhaps the most interesting program with a low introductory price is usage-proportional pricing. The remainder of this subsection investigates that plan under various assumptions.

We first consider the possibility that all subscribers have equal usage, proportional to the number of subscribers. It follows that price is also proportional to number of subscribers [instead of proportional to a quadratic function of number of subscribers as in the equilibrium model, equation (19)], and the expansion path of the service is a straight line.

One possibility would be to let $p = p_A \frac{q}{q_A}$. In that case demand would expand along OA . Price would increase automatically as q increased. When q reached the optimum, q_A , price would just equal its optimum, p_A .

An alternative is to let $p = p_A \frac{q}{q_B}$. Demand would expand along OB until the critical mass was reached. Then, for further expansion price should be fixed at $p_A(-\epsilon)$, and demand would expand along BA .

Unfortunately, the equal-usage assumption may not be very realistic. If not, usage-proportional pricing would exclude individuals with a lot of low value usage and would admit (for a low price) individuals with a small amount of high value usage.

This might not be a bad idea, even in the long run. In general, usage-proportional pricing would be appropriate if the network were being used at capacity and costs were closely related to total usage, not necessarily to the number of subscribers. A fixed price per subscriber would be appropriate if the network were not being used at capacity and much equipment had to be committed for each subscriber.

The latter is perhaps more typical of a new communications service. Thus, usage-proportional pricing would probably result in some inefficiencies and misallocation of resources. Whether these inefficiencies are substantial or trivial is an empirical question. In any event they should be compared to the inefficiencies of other start-up programs; e.g., the direct approach discussed in the previous subsection.

For a mature service the price should probably have a higher fixed component and a lower usage component. However, the externalities in consumption still have to be taken into account.¹⁴

A program with a low introductory price relies more on market processes than does the direct approach discussed in the previous subsection and does not depend so critically on the managers' judgment. In particular, the managers are not required to determine which individuals have the highest w_i . The individuals select themselves by choosing to subscribe to the service at the offered price.

D. COMMUNITY OF INTEREST GROUPS.

Community of interest groups may greatly reduce the practical difficulty of starting up the service. Maximum equilibrium demand may be achieved even if the initial user set is small—so long as that set exceeds the critical masses for some community of interest groups.

At the same time, community of interest groups place a greater burden on whatever procedure is used to select the initial users. If the initial user set is selected by managers, they must know what the community of interest groups are and decide how many individuals to select from each. In some circumstances, it would be optimal to select everyone from the same group; in other cases a more even spread would be optimal. In any event, the managers must make this choice, and the success of the program may be greatly influenced by how well they choose.

Community of interest groups also place a greater burden on the market process for programs involving a low introductory price. The expansion path

involves price and g_1, \dots, g_k (instead of q as in Figure 4). Efficiency requires that k different linear combinations of the g_i be concave downward. (The equivalent condition to γ 's being concave downward in Figure 4.) But we have only one control variable, p . Clearly, it may be impossible to satisfy all k concavity conditions.

Thus, *any* program based on a (single) low introductory price may be inefficient in the sense that some individuals may join at the low introductory price but later drop out as the price rises. If this problem is serious, the seller may find it advantageous (and perhaps necessary) to use discriminatory pricing to assure that only "permanent" users join.

E. FEW SUBSTANTIAL CONTACTS.

If an individual's demand depends on his principal contacts' being users, the start-up problem may be fundamentally different from that discussed above. It may be unnecessary for there to be hundreds or thousands of users before an individual demands the service. A user population of two or three may be self-sufficient—if they are the right two or three people.

These small self-sufficient sets do not necessarily promote further growth but they do allow a different kind of approach to starting up the service. The seller can begin by establishing small self-sufficient user sets. He can then gradually expand the size and number of these sets until the desired equilibrium user set is attained or until the service starts to grow by itself. The practicality of this method depends primarily on the size of the (minimum) self-sufficient user sets.

The smallest possible self-sufficient user sets consists [sic] of two mutual contacts for whom both v_{ij} and $v_{ji} \geq p$. If there are many such pairs, i, j in the population, the service will start up, expanding beyond the minimum equilibrium set, with little or no help. Selling the service to i and j requires only getting the two together. And they may organize themselves and agree to subscribe to the service (contrary to the disequilibrium adjustment process assumed above¹⁵). After i and j both subscribe to the service, they may attract other individuals, and the service can grow further.

In fact, such growth from self-sufficient sets of two probably accounts in large part for the success in starting up telephone service. Indeed, telephone made a substantial penetration of the market while it was entirely a private-line service.

However, even if the service is viable, its incremental utility may be insufficient for very many people to demand the service to communicate with a single principal contact. If this is the case, the start-up problem is more difficult, and we must deal with larger self-sufficient user sets.

A self-sufficient set of three mutual contacts might also be able to organize itself and have all three members agree to subscribe together. However, this

becomes progressively more difficult and unlikely as the size of the self-sufficient sets increases. It would be especially difficult if all the members of the set were not mutual contacts and no one knew all the other members of the set.

The seller might be able to gather data and determine self-sufficient use sets. He could then try to sell the service to everyone in such a set simultaneously. Naturally this is more difficult, the larger is the self-sufficient set. In fact, the difficulties of organizing even six to eight people and getting them all to agree to a joint purchasing decision may be far from trivial.

Nevertheless, the seller might do well to gather data on communications patterns and try to determine self-sufficient user sets. But it may be necessary to combine this with some other kind of start-up program. This could take the form of the direct approach or the low introductory price previously discussed. However, it is also possible to have a continual program in which each new subscriber is offered a low rate until the seller can connect a self-sufficient set of users that contains him. This kind of program would be effective if new users had high values of w_i but little community of interest with the current user set. ▼

REFERENCES

1. ARTLE, R. AND AVEROUS, C. "The Telephone Service as a Public Good: Static and Dynamic Aspects." *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1 (Spring 1973), pp. 89-100.
2. MARRIS, R. *The Economic Theory of "Managerial" Capitalism*. New York: Free Press, 1964.
3. SQUIRE, L. "Some Aspects of Optimal Pricing for Telecommunications." *The Bell Journal of Economics and Management Science*, Vol. 4, No. 2 (Autumn 1973), pp. 515-525.

1 See [1].

2 An earlier attempt to model interdependent demand is Marris [2]. Marris develops a general theory of demand for new products but does not consider communications in particular. The interdependent aspects of demand have a much different interpretation in his analysis than in this paper or in Artle and Averous [1] or Squire [3] (discussed later in the text). Also, irreversibility plays a much larger role in Marris' analysis. Nevertheless, Marris develops some of the same concepts used in this paper; e.g. critical mass.

3 In [3]

4 We can also define equilibrium user sets with respect to any given set of discriminatory prices; i.e., a set of users such that

$$q_i = q_i^D(p_i, q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n) \tag{9a}$$

for all i , where p_i is the price charged to individual i .

5 P. 22

6 P. 22

7 In [1] and [3], respectively.

8 P. 26

9 P. 22

10 P. 22

11 P. 22

12 Of course, this issue becomes much more complicated if we consider interrelationships with other markets and possible income redistribution.

13 P. 29

14 See Squire [3].

15 P. 22.