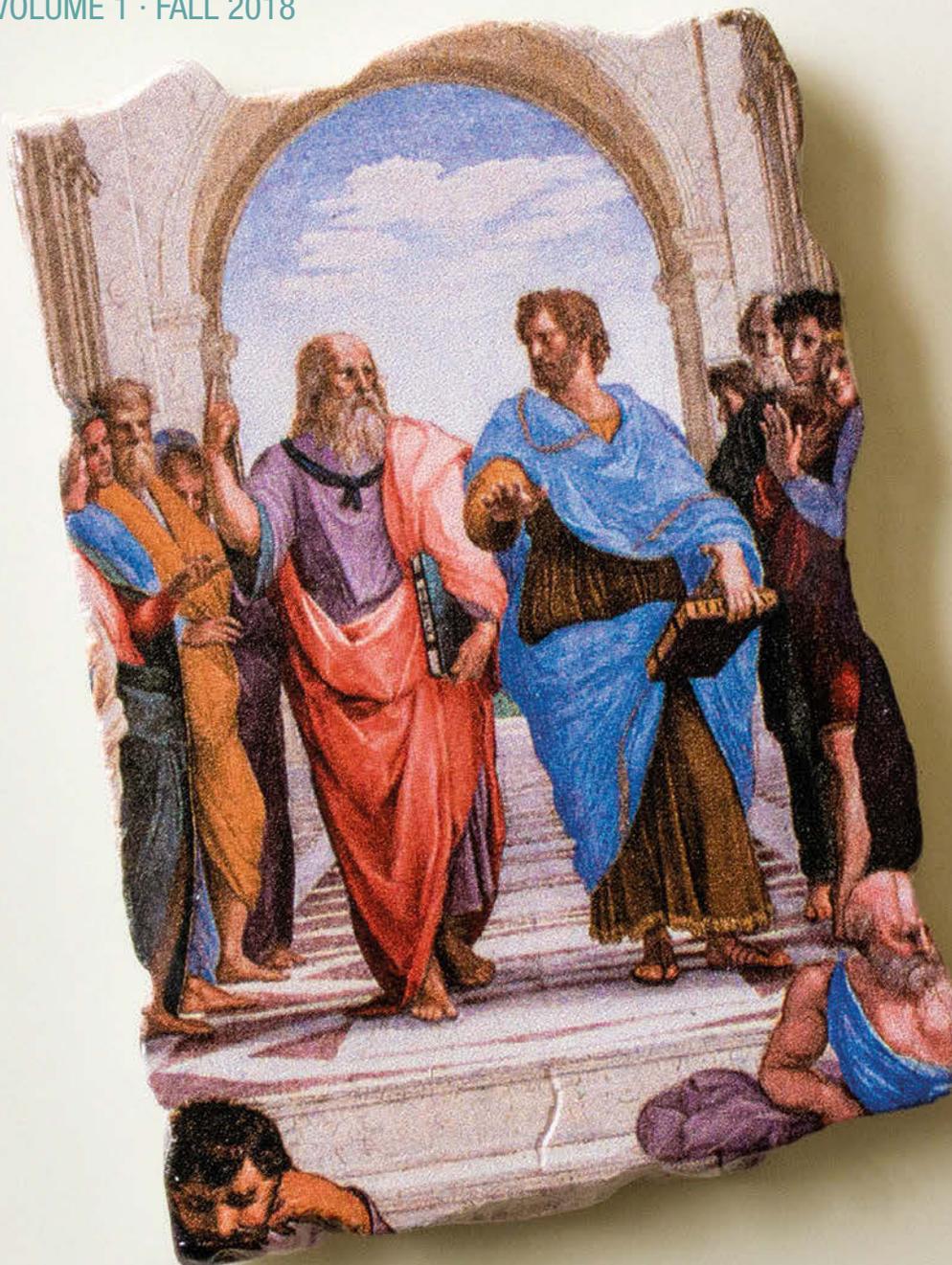


Antitrust Chronicle

OCTOBER · VOLUME 1 · FALL 2018



CRESSE Insights

TABLE OF CONTENTS

03

Letter from the Editor

22

Patent Reform, Innovation, and the Scope of Competition Policy
By Mark Schankerman & Florian Schuett

04

Summaries

28

What is the Scope for Choice and Competition in Education?
By Allan Fels & Dr. Darryl Biggar

06

What's Next? Announcements

35

Public and Private Antitrust Enforcement for Cartels: Should there be a Common Approach to Sanctioning Based on the Overcharge Rate?
By Yannis Katsoulacos, Evgenia Motchenkova & David Ulph

07

CPI Talks...
...with Frederic Jenny

45

A Competition Law Analysis of Common Shareholdings
By Neil Campbell

12

Two-Sided Red Herrings
By David S. Evans & Richard Schmalensee

54

Five not so Easy Pieces to Make Antitrust Work for Innovation
By Richard Gilbert

18

CRESSE: Actual and Potential Effects
By James S. Venit

Editorial Team

CEO & Founder

David S. Evans

Editor in Chief

Elisa V. Mariscal

Managing Director

Raúl Escalante

Managing Editor, CPI Projects

Elisa Ramundo

Managing Editor, The Antitrust Chronicle

Sam Sadden

Senior Editor

Nancy Hoch

Latin America Editor

Jan Roth

Junior Editor

Jeff Boyd

Editorial Advisory Board

Rosa Abrantes-Metz

Stern School of Business

Kent Bernard

Fordham School of Law

Rachel Brandenburger

Oxford University

Dennis W. Carlton

Booth School of Business

Adrian Emch

Hogan Lovells

Kyriakos Fountoukakos

Herbert Smith

Jay Himes

Labaton Sucharow

James Killick

White & Case

Stephen Kinsella

Sidley Austin

Ioannis Lianos

University College London

Robert O'Donoghue

Brick Court Chambers

Aaron Panner

Kellogg, Hansen, Todd, Figel & Frederick

Vanessa Yanhua Zhang

Renmin University

LETTER FROM THE EDITOR

Dear Readers,

The October 2018 CPI Antitrust Chronicle includes articles based on presentations from Special Policy Sessions (“SPS”) and invited lectures of the 13th Annual CRESSE Conference organized from June 29–July 1, 2018 in Crete, Greece.

CRESSE (www.cresse.info), is an international network of academics and other professionals, with an interest in Competition Policy and Sectoral Regulation. Every year CRESSE organizes an international conference in Greece that is widely recognized as one of the top academic conferences in the economics of competition policy and regulation worldwide. For the 2018 Conference, Keynote Speakers included John Vickers, Dennis Carlton, David Evans, Herbert Hovenkamp, and Eleanor Fox. An important feature of the annual CRESSE Conference is the organization of a number of SPS in which important topical issues of competition policy are discussed between academics (economists and lawyers), policy makers, corporate representatives, and practitioners.

The contributions here include articles by a number of prominent economists and legal experts. David S. Evans & Richard Schmalensee provide an account of the relevant economic modelling and of what sort of evidence is necessary or sufficient in markets with platform businesses to establish competitive effects, in the aftermath of the Supreme Court’s decision in *American Express*. James Venit discusses a potential flaw in the European Court of Justice’s *Intel* judgement as a result of the judgement focusing exclusively on the potential effects of the dominant firm’s conduct in an *ex-post* sense, where the allegedly abusive conduct has come to an end. Mark Schankerman and Florian Schuett in their contribution on patent reform, innovation and the scope of competition policy argue that many competition policy issues related to innovation stem from the “patent quality” problem – the prevalence of patents of dubious validity – and that fixing this problem at its source can reduce the scope for competition policy intervention. Yannis Katsoulacos, Evgenia Motchenkova, and David Ulph discuss recent theoretical literature pointing to the ineffectiveness in terms of welfare impact, of currently applied sanctioning regimes for cartels by Competition Authorities and argue the case for switching to a penalty regime, in which the penalty base continues to be the currently dominant penalty base of cartel revenue but in which, in contrast to current practice, the penalty rate is based on the cartel overcharge – which is commonly estimated in order to calculate damages in private damage claims. Neil Campbell’s article on a competition law analysis of common shareholdings argues that the application of merger control laws to mergers involving two companies in an industry with significant common shareholdings or to the acquisition of common shareholding positions by institutional investors is unlikely to be effective and the costs would probably exceed the benefits. Allan Fels deals with the scope for choice and competition in schools finding that while there appears to be much scope for choice and competition in education, there are also some significant, often ignored limitations if a major goal of education policy is to promote “equality of opportunity.” Last, in an article that concentrates on how antitrust should handle highly innovative markets, Richard Gilbert, CRESSE Keynote Speaker in 2019, discusses whether and how the assessment of potentially anticompetitive conduct should be amended when enforcing competition law in such markets.

Finally the October Antitrust Chronicle presents an interview with Prof. Frederic Jenny focusing on his thought-provoking lecture in CRESSE 2018 on the role of “fairness” as a goal of competition policy. He argues that competition authorities should care about fairness but at the moment competition law enforcement seems to have several weaknesses in this respect. Competition authorities rarely, if ever, look at the consequences of competition or anticompetitive practices on labor markets. Furthermore, on product markets competition authorities generally refuse to consider exploitative abuses of dominance such as charging high prices as a competition law violation. Equally they are reticent to deal with claims of “abuses of buying power.” In short, they refuse to pass judgment on the fairness of the results of negotiations between suppliers and consumers on competitive markets.

Yannis Katsoulacos

CRESSE & Athens University of Economics and Business

SUMMARIES

07



CPI Talks...

...with Frederic Jenny

In this month's edition of CPI Talks... we have the pleasure of speaking with Frederic Jenny. Professor Jenny is Chairman of the OECD Competition Committee. In this interview, he reflects on his lecture at CRESSE 2018 which was titled "Should competition authorities care about fairness and if so how?"

12



Two-Sided Red Herrings

By David S. Evans & Richard Schmalensee

A surprising amount of debate leading up to the Supreme Court's decision in *American Express*, and the commentary following this landmark ruling, attempt to trivialize and marginalize the modern economic learning on multisided platforms. Despite these efforts the 2nd Circuit Court of Appeals and the Supreme Court ultimately embraced the economic literature on these business models. This article debunks five red herrings that have been floated in the debate: (1) the two sides are just complements, nothing new there; (2) everything is two-sided, or who's to know what's two-sided; (3) as industries mature two-sidedness goes away; (4) markets must be one sided since the services to the two sides aren't interchangeable; and (5) two-sided analysis "devastates" antitrust law. The Supreme Court's decision has raised a host of interesting issues, including how to deal with two-sided platform businesses that look different from American Express's credit-card platform and what sort of evidence is necessary or sufficient in markets with platform businesses to establish competitive effects. Like any Supreme Court decision, not every word was chosen as carefully as it might have been, and clarifications will be needed going forward. The large and evolving literature on two-sided platforms will prove helpful to sort that out and we anticipate that the courts will embrace this constructive approach.

18



CRESSE: Actual and Potential Effects

By James S. Venit

The Court of Justice's *Intel* judgment, which rejects the application of *per se* condemnation of rebates granted by dominant firms when conditioned on exclusivity, and which implicitly endorses the use of the "As Efficient Competitor" test is a welcome step forward and a victory for those in DG Competition who have advocated the adoption of an effects-based approach under Article 102. However, the Court's judgment is flawed to the extent that it focuses exclusively on potential effects in an *ex-post* case where the allegedly abusive conduct had come to an end so that actual affects could be determined.

22



Patent Reform, Innovation, and the Scope of Competition Policy

By Mark Schankerman & Florian Schuett

This paper argues that many competition policy issues related to innovation stem from the "patent quality" problem – the prevalence of patents of dubious validity – and that fixing this problem at its source can reduce the scope for competition policy intervention. We summarize recent research in which we develop an integrative framework of patent screening that includes patent office examination, fees, and validity challenges in the courts. Simulations of the model, calibrated on U.S. patent and litigation data, indicate that the patent-quality problem is indeed serious. Low-quality patent applications are prevalent and patent office screening does not weed them out effectively. We identify a number of potential patent reforms and report estimates of their likely welfare effects. Our analysis suggests that frontloading fees and introducing a post-grant patent office review would considerably improve patent screening. By raising patent quality and thus reducing uncertainty around their validity, patent reform can solve some of the problems that competition policy would otherwise need to address.

SUMMARIES

28



What is the Scope for Choice and Competition in Education?

By Allan Fels & Dr. Darryl Biggar

Is there scope for greater reliance on competitive market forces in the provision of education services? We argue that whilst there appears to be much scope for choice and competition in education, and certainly greater scope than exists at present in many countries, there are also some significant, often ignored, limitations on the scope for choice and competition if a major goal of education policy is to promote “equality of opportunity.”

35



Public and Private Antitrust Enforcement for Cartels: Should there be a Common Approach to Sanctioning Based on the Overcharge Rate?

By Yannis Katsoulacos, Evgenia Motchenkova & David Ulph

In this article we discuss recent theoretical literature pointing to the ineffectiveness, in terms of welfare impact, of currently applied sanctioning regimes for cartels by Competition Authorities (CAs). We then, for the first time, provide a comparison of different regimes taking also into account criteria other than welfare, related to ease of implementation and transparency. We argue the case for switching to a penalty regime, in which the penalty base continues to be the currently dominant penalty base of cartel revenue but in which, in contrast to current practice, the *penalty rate* is based on the cartel overcharge – which is commonly estimated in order to calculate damages in private damage claims.

45



A Competition Law Analysis of Common Shareholdings

By Neil Campbell

An emerging economics literature has raised concerns that “common shareholdings” by institutional investors in multiple public companies may give rise to soft competition and the exercise of market power in concentrated oligopolies. However, it would be a mistake to try to address such concerns using merger control regimes. While competition laws which prohibit “competitor agreements” may provide a basis for dealing with explicit coordination between competitors that is facilitated by common institutional shareholders, the application of merger control laws to mergers involving two companies in an industry with significant common shareholdings or to the acquisition of common shareholding positions by institutional investors is unlikely to be effective and the costs would probably exceed the benefits.

54



Five not so Easy Pieces to Make Antitrust Work for Innovation

By Richard Gilbert

Innovation is an important concern for merger policy in high technology industries. While the antitrust laws are sufficiently flexible to address issues in the R&D-intensive sectors of the economy, there are several practical obstacles to effective antitrust enforcement for innovation. These include: (i) limited economic theory and empirical evidence relating mergers to innovation effects; (ii) the historical importance of market definition in merger analysis; (iii) the standard of proof to establish antitrust harm; (iv) the treatment of efficiencies and appropriability; and (v) limited evidence on the effectiveness of merger remedies to restore innovation incentives.

WHAT'S NEXT?

Our November 2018 Chronicle will focus on issues related to **Due Process & Antitrust**.

ANNOUNCEMENTS

CPI wants to hear from our subscribers. In the remaining months of 2018, we will be reaching out to members of our community for your feedback and ideas. Let us know what you want (or don't want) to see, at: antitrustchronicle@competitionpolicyinternational.com.

CPI ANTITRUST CHRONICLE DECEMBER 2018

The December 2018 Chronicle will have articles that address issues related to **Multi-Sided Markets & Consumer Harm**.

Contributions to the Antitrust Chronicle are about 2,500 – 4,000 words long. They should be lightly cited and not be written as long law-review articles with many in-depth footnotes. As with all CPI publications, articles for the CPI Antitrust Chronicle should be written clearly and with the reader always in mind.

Interested authors should send their contributions to Sam Sadden (ssadden@competitionpolicyinternational.com) with the subject line “Antitrust Chronicle,” a short bio and picture(s) of the author(s).

The CPI Editorial Team will evaluate all submissions and will publish the best papers. Authors can submit papers in any topic related to competition and regulation, however, priority will be given to articles addressing the abovementioned topics. Co-authors are always welcome.





...With Frederic Jenny

In this month's edition of CPI Talks... we have the pleasure of speaking with Frederic Jenny. Professor Jenny is Chairman of the OECD Competition Committee.

Thank you, Professor Jenny, for sharing your time for this interview with CPI.

1. Your lecture at CRESSE was titled “Should competition authorities care about fairness and if so how?” What do you think is the correct answer? And what type(s) of “fairness” are you considering, procedural, political or substantive for instance?

I think that the correct answer is that competition authorities should care about fairness. A law or a policy can be successful in the long run only if it is understood and considered to be legitimate by the public. If it is misunderstood or considered to be illegitimate, at some point, the law will be abolished or fundamentally changed.

We are witnessing a growing rejection of globalization, trade liberalization, and economic competition which feeds the populist rhetoric on both sides of the Atlantic. International and domestic market competition are increasingly seen as leading to unfair results such as the loss of jobs for some who have no or little hope of finding another job or downward pressures on wages for others. In the U.S., and to a lesser extent in Europe, owners of capital and highly skilled labor benefit greatly from competition while low skill workers are hurt increasing the inequality of income and wealth and the feeling that a few profit from market competition to the detriment of many others who are victims. Competition law enforcement seems to be unresponsive to the perceived injustice of the results of the competitive market system and does not seem to decrease the rising economic inequality which is the foundation of the sense of unfairness. If anything, the actions of competition authorities to promote more competition seem to promote more inequality and unfairness. As a result, economic competition law enforcement comes under criticism from populists both on the right and on the left. The results of econometric studies by the highly respected Brussels based Think Tank Bruegel on the votes for the Brexit in the UK and for Trump in the U.S. show clearly that the major determinants for both votes were inequality of income, poverty, and unemployment at the local level.

The claim by competition authorities that competition increases consumer welfare does nothing to comfort those who have seen their economic situation take a turn for the worse because of economic competition. Competition authorities are used to arguing that gains from economic competition could be used to compensate the losers but they are also quick to point out that redistribution is not their mission. Yet we know that redistribution schemes do not work well and that the losers from competition are not compensated. The sense of unfairness of the results of the competitive process is not limited to those who lose out in the competition process. One thing we learned from behavioral economists such as Kahneman or Thaler is that a large fraction of individuals has some notion of fairness as an argument in their utility function and, to a certain extent, prefers fairer outcomes to larger gains. The sense of unfairness of the results of the competitive process has without doubt been reinforced by the disruptions due to the 2008 economic and financial crisis, the rapid development of international trade with China, and the digital revolution.

Economic fairness has a number of dimensions: horizontal fairness among competitors, horizontal fairness among consumers, vertical fairness between suppliers and consumers, procedural fairness. One could argue that competition law enforcement addresses the issue of horizontal unfairness among competitors by emphasizing the need to fight the creation of artificial barriers to entry preventing potential competitors from accessing the market. Similarly, one can argue that competition law enforcement addresses the issue of unfairness among consumers by fighting price discrimination. Finally, the work undertaken at the OECD and the ICN on developing due process good practices tends to address the issue of procedural fairness. But the work of behavioral economists like Kahneman or Thaler has largely focused on exploring the notion of vertical fairness between suppliers and demanders, in particular, but not exclusively, on labor markets. With respect to vertical fairness, com-

petition law enforcement seems to have several weaknesses. Competition authorities rarely, if ever, look at the consequences of competition or anticompetitive practices on labor markets. Furthermore, on product markets competition authorities generally refuse to consider exploitative abuses of dominance such as charging high prices as a competition law violation. Equally they are reticent to deal with claims of “abuses of buying power.” In short, they refuse to pass judgment on the fairness of the results of negotiations between suppliers and consumers on competitive markets. But behavioral economics tells us that the result of a transaction, even in a competitive environment, can be considered unfair by individuals and, conversely, that the result of an anticompetitive practice is not necessarily perceived to be unfair. It is easy to understand that workers who lose their jobs as a result of market competition or see their wages being reduced may not consider that the benefits of market competition in the form of lower price are worth the cost to them of competition.

2. What are some aspects of competition policy that need to be addressed in response to the challenges of globalization, a global economy, and bigger, more powerful companies?

The promoters of market competition, both domestic and international, do not address some of the fundamental findings of economic theory and occasionally make assumptions which are patently contrary to economic reality.

First, the Stolper–Samuelson theorem of international trade makes clear that, as long as trading countries do not specialize completely, in each trading country, the factors of production which are scarcer will suffer as trade barriers are eliminated and international competition increases. In relative terms, if we compare the EU and the U.S. to other countries with which they trade, capital and highly skilled labor are abundant in the U.S. and the EU and unskilled labor is scarce. This means that capitalists and highly skilled workers will benefit from globalization while low skilled workers will be made worse-off. This distributional effect, which is well-known by economists, is rarely publicly discussed when trade liberalization measures are considered and appropriate remedies are even more rarely designed to accompany these trade liberalization measures.

The reason for this lack of interest in the distributional effects of international competition rests with the commonly held assumption by competition experts that reallocation of resources will take care of the problem of displaced labor. If some firms are unable to withstand competition, they will exit the market and their factors of production will be reemployed in other activities. If the labor market is a vast, competitive and undifferentiated market, this reallocation of resources will allow labor to find equivalent jobs elsewhere at roughly the same wage as the one they had prior to the reallocation of factors, which means that labor will not lose either in terms of employment opportunities or in terms of wages. The problem is that, in reality, the reallocation of labor does not take place because, contrary to the assumption made, the labor market is very fragmented both regionally and by technical skills. Labor, and particularly low skilled workers largely lack regional mobility for a number of reasons such as the fear that the other wage earners in the family will not be able to find a job in another location, or the fact that they have sunk money in a house and face a possible loss if they have to sell their house at a low price in an economically depressed area and have to find one at a high price in an economically booming area. Studies show that labor mobility in the U.S. has been on a downward trend since the 1990s. But labor is also not mobile technically. The programs to train unskilled workers to acquire usable skills in fast growing, often hi-tech, industries have largely failed to deliver meaningful results in most countries for lack of adequate financing and because of poor educational quality or an inability to match supply of skills and demand for skills.

If labor is not mobile then in industries subject to a competition shock, unskilled workers are likely to lose their jobs without any prospect of finding another one or likely to have to accept deep cuts in wages.

Finally, competition economists assume that consumers derive utility from the consumption of goods and services exclusively. But behavioral economics has shown that they care about fairness and are willing to trade fairness for consumption in order to increase their utility.

There are a number of ways for competition authorities to deal with the issue of fairness which do not imply abandoning the consumer welfare goal which has been central to competition law enforcement over the last thirty years or the adoption of a public interest goal in addition to the consumer welfare goal. Indeed, I do not believe that the populists are critical of the attempt to promote competition in order to maximize consumer welfare as such but rather critical of the unfairness of the result of the competitive process in certain circumstances, namely when a part of the labor factor is not mobile.

First, competition authorities could use their advocacy powers to advocate for policies which could facilitate the geographical and technical mobility of labor. By advocating for measures which could facilitate the reallocation of labor, competition authorities would advocate for the realization of conditions which could make competition work better and be perceived as more tolerable and fair. But up to now, competition authorities have not usually used their advocacy power to address the functioning of labor markets.

Second, among the cases of suspected anticompetitive practices, competition authorities could use the lessons of behavioral economics to prioritize the cases which will appeal to the sense of fairness of consumers. For example, Kahneman and Thaler show that most individuals (in the U.S.) find that an increase in price not justified by an increase in cost is unfair while most individuals tend to see an increase in price by a supplier to pass on an increase in cost as fair. Competition authorities could use such results to prioritize the cartels in which firms have increased their price without having experienced an increase in cost rather than crisis cartels or cartels formed to pass on to consumers an increase in the cost of suppliers.

Third, competition authorities could analyze the impact of the remedies they impose in merger control or in antitrust cases on the labor market because they cannot assume that these measures will never affect labor. If a merger is found to be pro-competitive but is likely to create a large disruption for labor (with substantial lay-offs and little prospect for the laid-off employees to find alternative employment), competition authorities could request the merging parties to take at least transitory measures to alleviate the labor disruption. This is to a certain extent what the South African competition law mandates.

3. Is there a need for a supranational competition authority dealing with cooperation and enforcement on a global scale?

The creation of a supranational authority requires by definition an abandonment of national sovereignty to this authority by the founding members of the supranational authority. In a world in which there are about 200 countries of different sizes, with different levels of technological development, different economic histories, different legal systems, etc... and at a time of resurgence of political and economic populism in some of the most important economic blocs in the world, it is clear that the idea of the creation of a supranational competition authority dealing with competition enforcement on a global scale is dead on arrival.

However, regional supranational authorities may be an effective way for sets of neighboring small countries to overcome some of the difficulties they have in dealing with transnational competition law issues raised by global world players. The example of Comesa in Africa, a regional grouping of 19 Member States in Eastern and Southern Africa, in which the Member States have established a central merger control mechanism is an interesting one.

At the global level there are two ways to make competition law enforcement more effective and less costly to businesses in a world characterized by a combination of economic globalization and fragmentation of nation-states: convergence of national competition law regimes and cooperation in competition law enforcement. Over the last thirty years we have made considerable progress on convergence in merger control and in the treatment of anticompetitive agreements (even though there are still some divergences with respect to abuse of dominance). There is still work to be done for convergence on due process issues but this work is undertaken in the context of the OECD and ICN.

With respect to international cooperation, voluntary bilateral and regional cooperation both informal and formal between competition authorities has developed rapidly, particularly but not exclusively in the merger area, and some new instruments, such as waivers, have facilitated the development of fruitful exchanges among competition authorities. However, international cooperation on competition law enforcement is still spotty, costly, and insufficient. The main limitations are due to differences in legal standards, the existence of limits on what information competition authorities can lawfully exchange, the low willingness of competition authorities in certain jurisdictions to cooperate with other competition authorities, the occasional absence of waivers given by investigated firms, a lack of resources or time preventing each competition authority from cooperating with many others, etc...

What is necessary is a move forward to find ways to systematize international cooperation, to make it more inclusive and to make it less costly for competition authorities. In all likelihood, this requires inventing new instruments of cooperation. What these instruments could be is difficult to say at this point. But work is being undertaken at the OECD to investigate several possibilities such as the development of international standards for comity, systems of mutual recognition of other agencies' decisions, or possibilities of deference to one "lead authority."

4. In a time of increasing nationalism in the developed world, how do you see the most dynamic developing countries (BRICS) outweigh the need for competitive and open markets while dealing with issues arising from industrial policy and globalized economies?

The question of the compatibility between competition policy and industrial policy is neither new nor limited to developing countries. It is a question, however that may be of particular interest for developing countries which are faced with the challenges of under-development. The question for these countries is how to spur economic growth in order to catch up economically and eradicate poverty.

Industrial policies that are horizontal, such as government funding basic research, education, and national infrastructure projects are usually not in conflict with competition. Selective measures, particularly measures aiming to give a competitive advantage to some domestic firms over others or over foreign suppliers, may be in conflict with competition objectives. However, if selective measures are adopted in order to try to correct a market failure, for example, by supporting firms which choose to invest in a certain industry or technology having positive externalities, they can be compatible with competition policy. So, we should not be dogmatic about industrial policy.

Where the risk is greatest is when developing countries try to boost economic growth by creating national champions in the hope that these firms will eventually become world-class competitors. The hope is often that offering protection to those firms will allow them to operate at a bigger scale and to go down the learning curve faster than they would otherwise and therefore allow them to become more efficient than they would through the process of competition. Such strategies succeed very rarely not only because government are not particularly good at picking winners but also, and perhaps more importantly, because they rest on the hypothesis that the incentives of the protected national champion will remain the same as if it were not protected. But economics has taught us that protection weakens the incentive to perform so that dynamic efficiencies may not materialize or may be much more limited than anticipated. In addition, and by design, those policies involve a trade-off between certain static efficiency losses due to diminished competition and hypothetical dynamic efficiency.

The adoption in all the countries of a competitive neutrality framework protecting producers from distortions created when rival firms receive subsidies or other support from their governments would be extremely useful to complement competition policy and to minimize the possible contradictions between competition policy and law enforcement, on the one hand, and industrial policy, on the other hand. Such frameworks unfortunately only exist in a handful of countries such as Australia, Sweden, Ukraine, the European Union, and China. But in the European Union, it is limited to state aids whereas industrial policy can take many forms, such as political involvement in strategic deals, joint technological or industrial initiatives, public-private partnerships for infrastructure, administrative hardship or ease on certain industries, licensing requirements, etc. . . . And in China it only applies to industrial policy measures at the provincial level but not to measures taken by the central government. The OECD Competition Committee is working on assessing the potential benefits from the adoption of competitive neutrality frameworks and on the design of such frameworks.

5. Is there a role for multiple goals for competition policy? Especially in light of the aims of some BRICS countries and other smaller competition authorities.

and

6. Altering the consumer welfare standard likely requires trading away some amount of consumer welfare in favor of some other values. How can antitrust authorities find the right balance, define the right goals, and use the proper tools for achieving them?

The enactment of a law is necessarily a political act. Because of this, competition laws often reflect the political realities of the countries in which they are adopted. In many countries, the protection of consumer welfare is seen as an insufficient reason to warrant the adoption of a competition law. Rather, it is considered that competition law should, while protecting economic competition, also promote the larger values of society. For example, a fundamental principle of competition policy and law in South Africa is the need to balance economic efficiency with socio-economic equity and development, which is understandable given the history of apartheid in South Africa and the extensive poverty in this country. One of the objectives of the Chinese anti-monopoly law, besides the enhancement of economic efficiency, is the promotion of the healthy development of the socialist market economy. The fact that competition laws may have public interest goals is not specific to developing countries. To take a simple example, the official goal of the EU competition law is to prevent competition from being distorted to allow the development of the internal, single, market.

It is generally considered that in countries in which there is a sufficient political consensus on the benefits of economic competition, it is preferable to assign only one goal to competition law, the promotion of consumer or total welfare, and to leave to other institutions the role of promoting the wider socio-political goals pursued (inclusiveness, independence, employment, etc. . .). The obvious reason for this stand is that the competition authorities are not well-placed, do not have the analytical tools, to trade-off efficiency benefits against other public interest costs (such as, for example, job losses). However, the logical implication of this stand is that transactions or behaviors will be allowed only if they meet both the efficiency test (implemented by the competition authority) and the public interest test (implemented by another administrative body). The risk is then that some transactions or behaviors will not be allowed, in spite of the fact that they have clear efficiency benefits, because they do not meet the public benefit test.

But in countries where competition law is only acceptable if it pursues other socio-economic or socio-political goals as well as economic efficiency, the alternative is either not to have a competition law at all or to have a competition law which has several goals (presumably an efficiency goal and one or several public interest goals).

Firms which are operating on the global market, under certain conditions, prefer that countries to which they export or in which they are considering investing have a competition law framework rather than not having one, even if the competition law includes some public interest goals. Indeed, the existence of such a competition law framework is seen as a protection against possible anticompetitive strategies by the domestic firms of the country receiving exports or in which investments could be made and the lack of such a framework as a threat that the “law of the jungle” will prevail.

However, international businesses tend to expect that four conditions should be met in countries which have public interest goals in their competition law. First, the public interest goals must be clearly and precisely specified in the law; second, the competition analysis of the behaviors or transactions reviewed by the competition authority should be separate from the public interest analysis of those behaviors or transactions; third, the competition authority should publish guidelines about its interpretation of the public interest tests it is required to do; and fourth, the balancing between the efficiency effects of the transaction or the behavior analyzed by the competition authority and its public interest effects must be transparent.

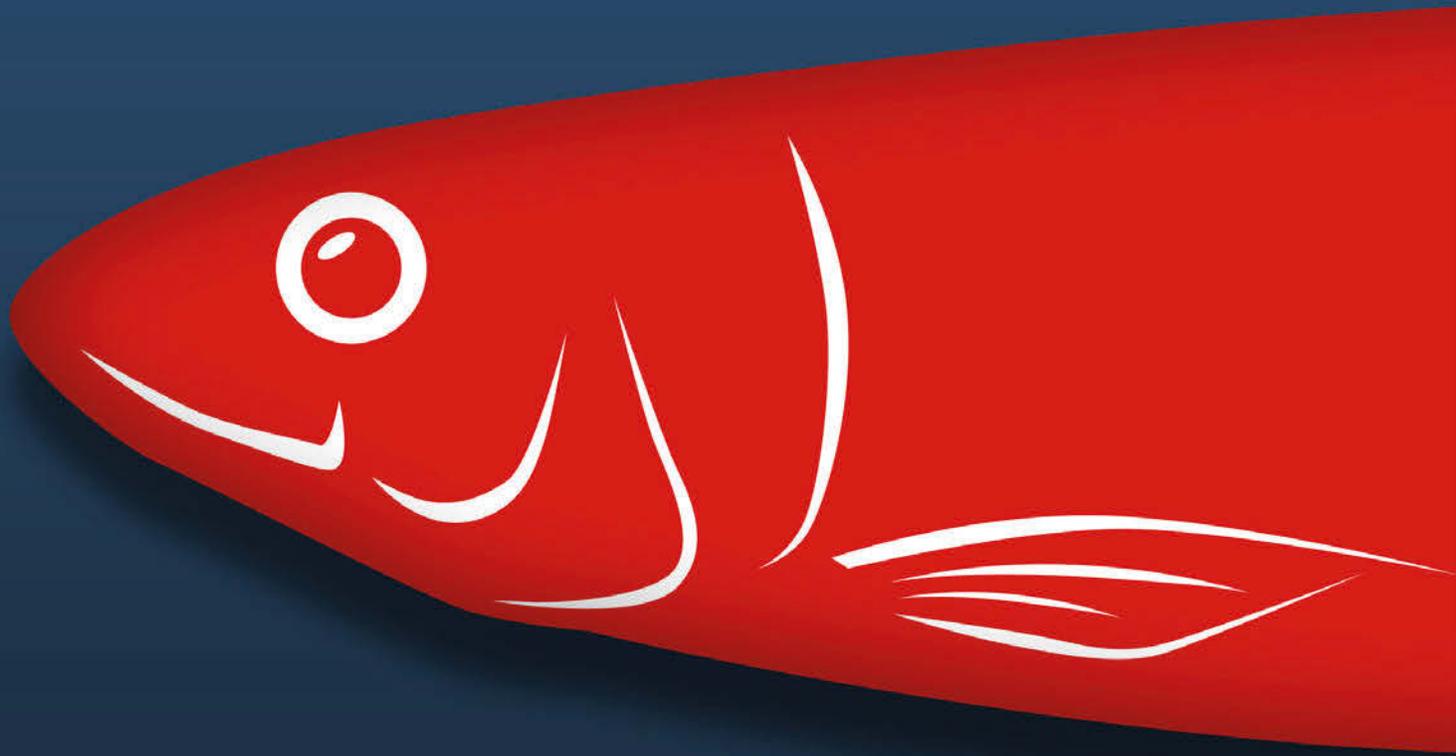
This pragmatic position rests on two ideas; first, the idea that competition authorities, even when the competition law they enforce only has an efficiency goal, already have to make difficult trade-offs (for example, trading off a risk of price increase against a possibility of innovation in a merger control); and, second, the idea that, from the standpoint of private businesses, legal predictability (ensured by transparency of the process) is more important than the economic optimality of the competition law enforcement.

To a very large extent, it is precisely because the process implemented by the South African Competition Authority meets these conditions that it is considered to be tolerable.

In a number of countries, as the public and decision makers have become more convinced of the benefits of using economic competition to promote consumer or total welfare, the public interest goals which may be written into competition law either are eliminated by amending the law or are given less weight in the interpretation of competition law which then begins to converge with economic reasoning. This latter evolution has characterized the EU and its Member States since the beginning of the 21st century. History suggests that this evolution often happens several decades after the adoption of a competition law and that, therefore, we should be patient with countries which have only recently enacted a competition law with public interest goals. Effective advocacy on the part of competition authorities is key to speeding up this evolution.



TWO-SIDED RED HERRINGS



BY DAVID S. EVANS & RICHARD SCHMALENSEE¹



¹ Evans is Chairman, Global Economics Group, Boston MA; and Executive Director, Jevons Institute for Competition Law and Economics; and Visiting Professor, University College London, London, UK. Schmalensee is Dean Emeritus and Howard W. Johnson Professor of Management Emeritus at the Massachusetts Institute of Technology (“MIT”) Sloan School of Management and Professor of Economics Emeritus at the MIT Department of Economics.

I. INTRODUCTION

The Supreme Court's recent *American Express* (“AmEx”) decision has raised a host of interesting issues, including how to deal with two-sided platform businesses that look different from AmEx's credit-card platform and what sort of evidence is necessary or sufficient in markets with platform businesses to establish competitive effects.² The large and growing economics and business strategy literature on two-sided platforms, now almost two decades old,³ will be helpful in sorting out these and other issues, as the extensive citations to that literature by the District Court, Appeals Court, and Supreme Court indicate.⁴

Unfortunately, a considerable amount of the recent debate in the U.S. on how to conduct antitrust analysis of two-sided businesses has involved attempts to trivialize or marginalize the findings of the relevant economics literature. This is surprising because there have been no critical comments on the main papers in this literature, which have appeared in leading economics journals beginning in 2003. A co-author of the seminal paper on two-sided platforms was awarded the 2014 Nobel Prize in Economic Science for a body of work that included this subject.⁵ In the *AmEx* litigation, the District Court and Appeals Court both cited this literature without any criticism.

In what follows we discuss five red herrings — assertions that have been used to marginalize the role of the extensive economics learning on two-sided platforms in antitrust analysis.

II. RED HERRING 1: IT'S JUST ABOUT COMPLEMENTS, LIKE GASOLINE AND TIRES

The first red herring says that there is really nothing novel about two-sided platforms because the services on the two sides are just complements, and the courts know what to do (and not to do) with complements.⁶

In order to process transactions between merchants and cardholders, American Express must provide services to merchants and different services to cardholders, and the prices to either group will affect that group's participation on the platform and thus the attractiveness of the platform to the other group. In oral argument, Justice Breyer contended that this interdependence in demand is simply what characterizes complements, like nuts and bolts.⁷ In his dissenting opinion, he compared the different services to gasoline and tires.⁸ In an amicus brief, eight economists compared them to tennis balls and tennis racquets.⁹

This argument has no merit. It does not appear anywhere in the serious economics literature on two-sided platforms. None of the economics journals that have published the key theoretical articles have published critical responses that say that the theory of two-sided platforms is retreading well-known concepts about complements. The claim is simply wrong.

2 *Ohio v. American Express Co.*, 138 S.Ct. 2274 (2018). Some platforms have more than two sides, and all that we say here also applies to such multisided platforms.

3 That literature began around 2000 with circulation of working paper versions of Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. EUR. ECON. ASS'N 4, 990-1029 (2003). Other key contributions include Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37 RAND J. ECON. 3, 645-667 (2006); Mark Armstrong, *Competition in Two-Sided Markets*, 37 RAND J. ECON. 3, 668-691 (2006); and E. Glen Weyl, *A Price Theory of Multi-Sided Platforms*, 100 AM. ECON. REV. 4, 1642-1672 (2010). There are no articles in serious economics journals, including the ones that published these papers, that argue that the theories described in these foundational pieces are wrong. For nontechnical surveys, see generally Marc Rysman, *The Economics of Two-Sided Markets*, 23 J. ECON. PERSPECTIVES 3, 125-143 (2009) and David S. Evans & Richard Schmalensee, *The Antitrust Analysis of Multisided Platform Businesses*, THE OXFORD HANDBOOK INT'L ANTITRUST ECON., VOL. 1 (Roger D. Blair & D. Daniel Sokol eds., 2014). The online appendix to the latter paper lists over 350 significant economics articles published through December 2012, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2185373. The Harvard Business Review has been publishing articles for managers on two-sided platforms since 2006. See Thomas Eisenmann, Geoffrey Parker & Marshall W. Van Alstyne, *Strategies for Two-Sided Markets*, 84 HARVARD BUS. REV. 92 (2006).

4 *Supra* note 2, *U.S. v. American Exp. Co.*, 88 F.Supp.3d 143, 165 (E.D. N.Y. 2015), and *U.S. v. American Exp. Co.*, 838 F.3d 179 (2d Cir. 2015).

5 Press Release, The Nobel Prize, The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2014 (October 13, 2014), <https://www.nobelprize.org/prizes/economics/2014/press-release/>.

6 For a related discussion see Lapo Filistrucchi, *Complements vs. Two-Sided Markets*, CPI ANTITRUST CHRONICLE, (Sept. 2018).

7 *Ohio v. American Express Co.*, (No.16-1454) (Oral argument February 26, 2018) at 22-24.

8 *Ohio v. American Express Co.*, 585 U.S. 11-12 (2018) (Breyer J., dissenting).

9 Brief for John M. Connor et al. as Amici Curiae Supporting Petitioners, *Ohio v. American Express Co.*, (No. 16-1454) (hereinafter Economists' Brief) at 4.

Two complements are usually both sold to the same customers; that's the reason why the price charged for one of the products affects the demand for the other. In contrast, American Express provides merchant and consumer services to members of distinct customer groups. The price charged to one side of the AmEx platform affects demand on the other side because of indirect network effects: merchants care about how many consumers use the card, and consumers care about how many merchants accept it.

Two complements can and often are sold by different firms. Many convenience stores sell tennis balls without selling tennis racquets, and companies often specialize in selling gasoline or tires but not both. By contrast, AmEx must serve both merchants and consumers to stay in business and must do so essentially simultaneously.

Finally, two-sided businesses like AmEx always facilitate interactions between customers on both sides of the platform. But you can buy gasoline without having any interaction with anyone who bought a tire.

III. RED HERRING 2: IF ANY BUSINESS IS TWO-SIDED, SO ARE ALMOST ALL BUSINESSES

The expansive variant of this red herring says that all businesses deal with members of more than one group — retailers, for instance, deal with both suppliers and customers — and therefore antitrust defendants will claim their business is two-sided. According to Professor Sagers, one of the signers of the amicus brief by lawyers in support of the Petitioner,

... we can expect every antitrust defendant and their sister to start claiming their business is two-sided, and lower courts will find reason within the theory to give their claims the time of day. After all, even a brick-and-mortar retail store is “two-sided” in the sense that it must balance the demands of suppliers and customers.¹⁰

Katz and Sallet provide a lighter version of this red herring. They claim there is a “lack of consensus regarding the definition of a platform [and] it is much harder to distinguish single-sided businesses from multisided ones than one might initially expect[.]”¹¹ They conclude that,

[g]iven the lack of definitional consensus regarding multisided platforms, coupled with the prospective applicability of existing definitions to a vast range of firms, it would be a mistake for antitrust enforcement to dramatically differ based on the threshold, and easily manipulable, question of whether a defendant is classified as a multisided platform.¹²

This suggests that there is an ongoing debate about the definition of two-sidedness. But there is no such debate. We believe that most, if not all, economists who have worked in this area would consider businesses to be two-sided platforms when there are (a) indirect network effects between members of at least one of the two customer groups and members of the other group; (b) these indirect network effects are strong enough to affect business conduct; and (c) the platform facilitates interactions between members of the two groups.^{13, 14} Moreover, the serious economics literature often identifies the same narrow set of businesses, defined by these characteristics, as two-sided. There are other nuances involving two-sided platforms. People who write articles in this area, including us, often include an abbreviated definition that doesn't go into all of these characteristics, which of course doesn't mean that they aren't recognized.

10 Chris Sagers, *Ohio v. American Express: Clarence Thomas Sets Sail on a Sea of Doubt, and, Mirabile Dictu, It's Still a Bad Idea*, ProMarket (June 27, 2018), <https://promarket.org/ohio-v-american-express-clarence-thomas-sets-sail-sea-doubt-mirabile-dictu-still-bad-idea/>.

11 Michael Katz & Jonathan Sallet, *Multisided Platforms and Antitrust Enforcement*, 127 THE YALE L REV. 7, 2142-2175 (2018) at 2148.

12 *Id.* at 2152.

13 After suggesting that there is an ongoing debate, Katz & Sallet assert that a good approach for antitrust purposes “is to define a firm as a multisided platform when *cross-platform network effects* occur in at least one direction and the firm facilitates interactions between two or more groups of users, can set distinct prices to different user groups, and has market power with respect to these groups.” See Michael Katz & Jonathan Sallet, *Multisided Platforms and Antitrust Enforcement*, 127 THE YALE L REV. 7, 2142-2175 (2018) at 2150. With the exception of the market power requirement this definition is, in our experience, consistent with how economists working in this area define two-sided platforms in practice. And Katz & Sallet, note 26, observe “Our inclusion of market power is meant to capture the likely circumstances in which antitrust issues arise, not to suggest that all firms with multisided business models have market power.”

14 In his dissent, Justice Breyer notes that Rochet & Tirole (2006, p. 646) also specify that if a business is a two-sided platform, the volume of transactions it manages “depends on the structure and not only the overall level of fees charged.” Like Katz & Sallet, we tend not to stress this condition, as it is an implication of indirect network effects in the absence of arbitrage between the two sides that would make it impossible for the platform actually to determine the effective price structure. As Rochet & Tirole (*id.*) put it, the structure of charges “is relevant only if the two sides do not negotiate away the corresponding usage and membership externalities.” We are not aware of evidence that platform businesses cannot use the price structure to balance demand because of arbitrage which would in any event be difficult given that platforms can often charge membership as well as transaction fees.

The definition commonly used by economists is narrow enough to exclude most ordinary, one-sided businesses. Supermarkets deal with both customers and suppliers, for instance. And supermarket customers may care about the variety and quality of goods on offer, but suppliers to supermarkets generally care only about their sales, not the number or characteristics of shoppers. More critically, supermarket customers and supermarket suppliers do not interact. Anchor Steam does not know that one of us has just bought a six-pack of their beer, just that a six-pack has been sold. There may be businesses for which this definition does not yield a clear conclusion, but they seem to be rare, and detailed inquiry into the facts of real businesses will usually resolve the issue.

The vast sea of doubt and uncertainty portrayed by some commentators on the definition of two-sidedness doesn't correspond to the now vast economics literature on this topic. The courts should not get sucked into a "what about" argument that makes a spurious claim that some business is two-sided — for example, "brick-and-mortar stores" — using a definition that isn't employed in the economics literature to avoid serious analysis of businesses that are likely to be two-sided based on a widely accepted definition and analytical methods.¹⁵

Justice Breyer says under the *AmEx* majority's definition, two-sided platforms are "commonplace."¹⁶ While the majority's definition is arguably less narrow than the one advanced above, the economics literature has found that many businesses are in fact two-sided, and it is widely recognized that this model has become more important as a result of the Internet and other related technologies.¹⁷ That is why two-sided platforms have attracted increasing attention among competition authorities around the world.¹⁸

We agree that some definitions of two-sidedness that have been advanced outside the economics literature are so broad as to imply that almost any business can be described as a two-sided platform. There is some merit to Justice Breyer's criticism that the definition employed by the *AmEx* majority was incomplete.¹⁹ But that's an argument for tightening up definitions in antitrust analysis, not throwing out an uncontroversial body of economics learning, and ignoring substantial cross-side effects when the facts show they are important.

IV. RED HERRING 3: TWO-SIDEDNESS IS IRRELEVANT IN MATURE MARKETS

This two-part red herring applies to mature markets with two-sided platforms, one in which all or almost all potential customers are engaged with one or more platforms. It is first argued that as a market matures, indirect network effects at the market level weaken and ultimately vanish when the market is fully mature. It is then argued that this change implies that indirect network effects at the firm level also weaken with market maturity and vanish in fully mature markets. Thus, even if firms were two-sided platforms before their market matured, once it has matured, it is argued, the links between the demands on their two sides have vanished.

To our knowledge this two-part argument was first made in expert testimony in the *Sabre* case and was accepted by the trial court.²⁰ It was repeated in an amicus brief to the Supreme Court in the *AmEx* case.²¹ In the economics literature it has only appeared as an assertion, without theoretical or empirical support, in a three-sentence paragraph in a single article.²²

15 Chris Sagers, *Ohio v. American Express: Clarence Thomas Sets Sail on a Sea of Doubt, and, Mirabile Dictu, It's Still a Bad Idea*, ProMarket (June 27, 2018), <https://promarket.org/ohio-v-american-express-clarence-thomas-sets-sail-sea-doubt-mirabile-dictu-still-bad-idea/>.

16 *Ohio v. American Express Co.*, 585 U.S. 17 (2018) (Breyer J., dissenting).

17 David S. Evans & Richard Schmalensee, *Matchmakers: The New Economics of Multisided Platforms* (2016); Geoffrey G. Parker, Marshall W. Van Alstyne & Sangeet Paul Choudary, *Platform Revolution: How Networked Markets Are Transforming the Economy - and How to Make Them Work for You* (2016).

18 OECD, *Rethinking Antitrust Tools for Multi-Sided Platforms* (2018), www.oecd.org/competition/rethinking-antitrust-tools-for-multi-sided-platforms.htm; Press Release, Federal Trade Commission, *FTC Announces Hearings On Competition and Consumer Protection in the 21st Century* (June 20, 2018), <https://www.ftc.gov/news-events/press-releases/2018/06/ftc-announces-hearings-competition-consumer-protection-21st>.

19 *Ohio v. American Express Co.*, 585 U.S. 15-18 (2018) (Breyer J., dissenting). Justice Breyer points out that the majority decision did not reference the point, made by Rochet & Tirole, that for a business to be two-sided the price structure must affect the overall volume of output. See Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 37 RAND J. ECON. 3, 645-667 (2006) at 664-665. Rochet & Tirole, being aware of their own definition, refer to credit cards as two-sided platforms in this paper as well as in their seminal 2003 publication. Also, we noted above, this role of the price structure is an implication of indirect network effects in the absence of arbitrage and is therefore a consequence of a business being two-sided rather than a defining characteristic.

20 *U.S. Airways, Inc. v. Sabre Holdings Corp. et al.*, 105 F. Supp. 3d 265 (2015).

21 Economists' Brief, *supra* note 9.

22 Oz Shy, *A Short Survey of Network Economics*, 38 REV. IND. ORG. 2, 119-149 (2011) at 136.

The argument that indirect network effects at the industry level generally weaken as a market matures is somewhat plausible, though we know of neither theoretical arguments nor empirical evidence that supports it. But even if it is true, it is not at all plausible that indirect network effects at the *firm level* are absent in mature markets. To see this, suppose there is a fixed number of possible participants on each side of a set of competing platforms and that all will join one or more platforms regardless of price or small changes in the number of participants on the other side of the platform. In this case, membership demand at the *market level* by each group is independent of the other group's demand at the margin, regardless of price, and indirect network effects at the margin are effectively absent at the market level.

Suppose further, to track the *Sabre* case, that one group of customers multi-homes by participating on all platforms (the airlines) and the other group of customers single-homes by participating on only one platform (the travel agents). The price charged by any individual platform to the single-homing group determines the extent to which travel agents join that platform versus competing platforms. That in turn determines how much the multi-homing group would pay for access to that platform. But even with perfectly fixed demand at the market level, if an *individual platform* charged travel agents too much (or subsidized them too little) it could lose all of those customers, and the airlines would have no reason to use that platform. That result is consistent with the effective absence of indirect network effects at the market level because, by assumption, all airlines and all travel agents would still join one of the remaining (competing) platforms. But, in this extreme hypothetical case, there would be one less platform competing in the market.

Thus, even if indirect network effects are weak or absent at the market level because the market has matured, there is no reason to think that they are weak or absent at the level of the individual firm. Making that assumption would be very likely to lead to an erroneous evaluation of individual firm conduct.

V. RED HERRING 4: LACK OF INTERCHANGEABILITY ON TWO SIDES IMPLIES TWO SEPARATE MARKETS

In an amicus brief to the Supreme Court in the *AmEx* case, a group of antitrust law professors presented an absurd market definition for two-sided platforms, proceeded to demolish it, and argued that their demolition proved that single-sided analysis was always appropriate.²³ They began by noting that the two groups served by two-sided platforms consume different services that are often not interchangeable. The services AmEx provides to merchants are clearly not good substitutes for the services it provides to consumers. They went on to argue that it would, accordingly, make no economic sense to include both sets of services in the same market, so that services to each group must be analyzed separately.

This is a very bright red herring. Consider competition among person-to-person money transfer services. It is true that the service provided to a person who sends money is literally different from, and not interchangeable with, the service provided to a person who receives money. Defining separate markets for sending money and receiving money, however, would ignore the core business reality that suppliers compete for transactions between senders and receivers. The transactions between senders and receivers *are* substitutable across competing money transfer platforms. An increase in the price of the transaction by one platform — almost no matter how that price is divided between the sender and receiver sides — would tend to result in an increase in demand for other platforms. Platforms that provide similar jointly consumed services are substitutes for each other, and their products are interchangeable as a matter of business reality.

Market definition should therefore focus on identifying suppliers that provide services that are interchangeable in this sense, which typically accords with business reality and sound economics.²⁴ The objective of market definition is to identify competitive constraints. Since the early 1980s the modern approach to market definition accordingly focuses on the ability of a firm or firms of interest to raise price above competitive levels.²⁵ It is not possible to make that assessment by looking at one side of a service that is consumed jointly by the two sides. The claim that one should exclude the other side of the transaction from the market because it isn't "interchangeable" is a red herring because it focuses on a service that the platform cannot provide separately and ignores the service that the platform provides jointly.

23 Brief of 28 Professors of Antitrust Law as Amici Curiae Supporting Petitioners, *Ohio v. American Express Co.*, (No. 16-154) at 17-20. Justice Breyer seems to accept this argument: *Ohio v. American Express Co.*, 585 U.S. 15-17 (2018) (Breyer J., dissenting).

24 Thus in the *AmEx* case, we have supported defining the market as consisting of payment services provided by AmEx and competing platforms. David S. Evans & Richard Schmalensee, *Applying the Rule of Reason to Two-Sided Platform Businesses*, 26 U MIAMI BUS L REV. 1, 1-15 (2018). Economists supporting Petitioners seem at times to agree with us on this point: they assert that "...the relevant competition occurs at the platform level (i.e. competition among the credit card companies)." Economists' Brief, *supra* note 9, p. 15.

25 See generally Gregory J. Werden, *The 1982 Merger Guidelines and the Ascent of the Hypothetical Monopolist Paradigm*, 71 ANTITRUST L.J. 1, 253-275 (2003); Dennis W. Carlton, *Market Definition: Use and Abuse*, 3 COMPETITION POL'Y INT'L 3 (2007); Carl Shapiro, *The 2010 Horizontal Merger Guidelines: From Hedgehog to Fox in Forty Years*, 77 ANTITRUST L.J. 49, 701-759 (2010).

Let's be clear on why this red herring is very dangerous. Through the rhetorical sleight of hand that different sides of the transaction aren't interchangeable, we are led to exclude the other side of the jointly consumed transaction, and the business realities of jointly competing for both sides, from the analysis.²⁶ Luckily, American courts are skeptical of market definitions that do not accord with business realities, as the plaintiffs found in *AmEx*.

VI. RED HERRING 5: CONSIDERING TWO-SIDEDNESS EXPLICITLY WILL *DEVASTATE* ANTI-TRUST

The colorful language is from Professor Wu.²⁷ This is as much a red herring, a distraction from substance, as the other assertions we have examined here. There is simply no reason why accounting for the business realities of two-sided platforms and relying on uncontroversial economics learning is going to do anything other than help courts and competition authorities make better decisions. If it does, modern antitrust analysis has a bigger problem than dealing with two-sided platforms. One-sided analyses of two-sided platforms can result in false negative decisions in addition to false positive ones.²⁸

The *American Express* decision necessarily left many issues unresolved, but it has made it clear that future cases will need to take the economics of two-sided platforms seriously. This will improve the quality of antitrust decisions. Around the world there is constructive discussion on how to do that, driven by the growing importance of platform businesses and recognition that they are in fact different from traditional ones in important respects, and that modern economic learning can help competition authorities and courts properly enforce the antitrust laws for them.

26 Another approach, which may be superior in some settings not involving the provision of services that are jointly and unseverably consumed, is to define separate markets for the services to each group the platform of interest serves but to take due account of the linkages between them in analysis. The more appropriate approach depends on both the facts of the case and the question at issue. A general rule that would require defining separate markets and ignoring linkages between them, which the antitrust law professors seem to advocate, makes no economic sense.

27 Tim Wu, *The Supreme Court Devastates Antitrust Law*, The New York Times (June 26, 2018), <https://www.nytimes.com/2018/06/26/opinion/supreme-court-american-express.html>.

28 David S. Evans & Richard Schmalensee, *Ignoring Two-Sided Business Reality can also Hurt Plaintiffs*, CPI ANTITRUST CHRONICLE (Apr. 2018).

CRESSE: ACTUAL AND POTENTIAL EFFECTS



BY JAMES S. VENIT¹



¹ Partner, Dentons, Brussels. The author represented Intel in the proceedings before the Commission and the General Court. However, the views expressed herein are purely those of the author. They have not been discussed with or reviewed by Intel and are not intended to reflect the views of Intel.

I. INTRODUCTION

In antitrust law certain agreements or forms of conduct, e.g. price fixing, are treated as *per se* illegal. That means that there is a broad consensus that the agreement or conduct is so anti-competitive that there is no need to examine its effects or the context in which it occurs in order to establish its illegality. This approach makes sense provided the conduct in question is so clearly anti-competitive that there can be no hesitation in condemning it. In such cases, strict, *per se* rules that require only establishing the existence of the conduct can both conserve enforcement resources and send a strong deterrent message.

Prior to the judgment of the European Court of Justice in the *Intel* case,² the European Commission and the Court of First Instance³ took the position that, absent some overriding justification, exclusive rebates granted by a dominant firm were *per se* illegal.

In particular in its *Intel* judgment the General Court had taken the view that:

- it was not required to examine all the circumstances in order to assess whether Intel's exclusive rebates were likely to have a foreclosure effect;⁴
- there was no need to analyze actual effects or consumer harm to determine the anti-competitive effects of rebates conditioned on exclusivity or quasi-exclusivity, even in an *ex-post* case;⁵
- there was no need to establish a causal link between the alleged abuse and actual effects on the market,⁶ or between the abuse and consumer harm;⁷
- there is no *de minimis* defense under Article 102.⁸

The General Court also rejected the need for a price/cost test to determine whether the rebates had the potential to foreclose AMD.⁹

In its *Intel* judgment, the European Court of Justice rejected the General Court's approach and clarified EU law by ruling that, where the defendant submits evidence supporting a claim that its conduct was not capable of producing foreclosure effects,¹⁰ the Commission is required to analyze not only the extent of the undertaking's dominant position, but also the part of the market covered by the challenged practice, the conditions under which the rebates are granted, their duration and amount, and the existence of a strategy to exclude an "as efficient" competitor.¹¹ The Court further held that, if the Commission uses the As Efficient Competitor Test (the "AEC test") to assess the capacity of the rebates to foreclose an equally efficient rival, then the General Court "*must examine all of the applicant's arguments seeking to call into question the validity of the Commission's findings*" about the rebates' capability to foreclose.¹²

2 Judgment of September 6, 2017, *Intel v. Commission*, EU:C:2017:632

3 Judgment of June 12, 2014, *Intel v. Commission*, T-286/09, EU:T:2014:547

4 GC §§ 80-85.

5 *Ibid* § 103.

6 *Ibid* § 104.

7 *Ibid* § 105.

8 *Ibid* §§ 116-120.

9 *Ibid* §§ 142-151. In its decision the Commission had applied the "as efficient competitor" test (the "AEC test") to determine whether AMD could supply the contestable part of customers' demand above cost and stated that the results of the AEC test corroborated its findings that Intel's discounts were exclusionary. In its judgment the General Court rejected the relevance of the AEC test in cases of both exclusive and loyalty-inducing discounts and declined to examine whether the test – which had taken up some 150 pages of the Commission's decision – had been properly applied.

10 *Ibid* § 138.

11 *Ibid* § 139.

12 *Ibid* § 141.

The Court's approach represents a significant victory for those within DG Competition who have advocated for an effects-based approach to Article 102 and the use of the AEC test. It also serves as a case study on how not to bring about a fundamental change in the approach taken by a competition authority. If the Commission had used the AEC test to reject a complaint against a dominant firm and the Legal Service had then had to defend the Commission on appeal, it is likely that the Legal Service would have defended the use of the AEC test. Conversely, it is not that surprising that, in response to Intel's appeal, the Commission's Legal Service opted for a traditional line of defense which required only that it show that Intel had conditioned its rebates on exclusivity and which thus avoided both the need to show potential foreclosure effects and the complexities and uncertainties of the AEC test.

By clarifying the existing jurisprudence to require EU Courts and the Commission to examine the relevant circumstances in order to assess the likelihood of foreclosure even in those cases where rebates are conditioned on exclusivity (at least where the defendant has made credible arguments challenging the foreclosure effect of such rebates), the Court resolved the binary choice with which it was confronted, between the General Court which had rejected an effects-based approach and the Advocate General, who embraced it, in favor of the latter. Unfortunately, the Court did not go on to analyze any of the five factors it identified as being relevant to the analysis of the likelihood of foreclosure effects, ruling instead that the General Court's judgment should be annulled because it had failed to consider Intel's arguments challenging the Commission's application of the AEC test.¹³ Thus, the Court's judgment rests on a procedural error, which obviated the need to consider additional factors, and in particular the extent of market coverage, that the Court had itself identified as requiring analysis.

The Court of Justice's approach towards the key issue of whether the facts are relevant in the case of rebates conditioned on exclusivity can also be characterized as procedural. Unlike the Advocate General, who provided a well-reasoned argument for treating exclusive and loyalty-inducing rebates in the same manner, the Court avoided both this issue and any discussion of the *Hoffmann-La Roche* jurisprudence¹⁴ on which both the Commission and the General Court had relied to justify their *per se* approach. Rather, the Court's "clarification" of the existing case law seems to rest on a procedural point: where the defendant advances plausible arguments challenging the potential foreclosure effects of an exclusive rebate, the Commission and the Courts should examine them rather than refusing to consider them simply because the rebate was conditioned on exclusivity.

On its face, the Court's decision not to examine Intel's claim that 3,5 percent market coverage during the last two years of the infringement was insufficient for there to be any foreclosure effects (or to send this issue back to the General Court) appears odd. As in the case of Intel's claims about the misapplication of the AEC test, the General Court had not examined this issue – or rather had sought to avoid it by arguing that since Intel's conduct was part of a plan to exclude AMD, market coverage should be assessed over the entire infringement period (on the average 14 percent) and not for the last two years. This would seem to be the same type of procedural error as with the AEC test. However, it cannot be excluded that the Court's approach was influenced by its endorsement of the General Court's conclusion that Intel had engaged in a single continuous infringement because of its strategy to exclude AMD. In any event, by not referring the issue of market coverage back to the General Court, the Court avoided taking a direct position on whether there is a *de minimis* defense under Article 102, although its judgment would seem to suggest that such a defense does exist, notwithstanding the Court's rejection of such a defense in *Post Danmark II*.¹⁵

As can be seen from the foregoing discussion, the *Intel* judgment is something of a mixed bag. On the positive side, the judgment confirms that dominant firms may compete on the merits even if this results in the exclusion of less efficient competitors. Second, the Court has clarified that, even in the case of exclusive rebates, the Commission and the Courts need to examine the relevant circumstances to assess potential foreclosure effects, at least where the defendant has challenged the likelihood of foreclosure. Third, it reaffirms that the AEC test may be a useful tool in assessing the potential to foreclose, thus implicitly rejecting the General Court's conclusion that the AEC test is too lenient because it would permit conduct that makes market access more difficult. Fourth, the Court has identified five factors that should be examined in order to assess the potential to foreclose of a rebate conditioned on exclusivity, although it unfortunately declined to examine any of these factors in its judgment. Last, the Court breathed some life into the justification defense by noting that the same five factors to be assessed in determining potential foreclosing effects are also relevant for assessing whether rebates may be justified.

On the negative side, the Court did not directly address the most recent ECJ case rejecting a *de minimis* defense under Article 102.¹⁶ It also accepted the concept of a single continuous infringement where there is a strategy to foreclose, and declined to consider whether rebates that cover only between 24 percent and 42 percent of an OEM's demand can be characterized as requiring that OEM to purchase all or most of its needs from the dominant firm.

¹³ Ibid §§ 144 and 147–50.

¹⁴ Judgment of February 13, 1979, *Hoffmann-La Roche v. Commission*, 85/76, EU:C:1979:36.

¹⁵ Judgment of October 6, 2015, C-23/14, *Post Danmark A/S v. Konkurrenceradet*, EU:C:2015:651.

¹⁶ Ibid.

But most significantly, the *Intel* judgment remains anchored in a world of potential, as opposed to actual, effects. This is highly significant because Intel's alleged infringements had all been concluded before the adoption of the Commission's decision in 2009. Indeed, in the case of Dell, perhaps the most important OEM, the infringement period ran from 2002-2005. Moreover, in 2006, Dell shifted a significant portion of its x86 CPU demand to AMD. That it had done so was a well-known fact before July 26, 2007 when the Commission issued its Statement of Objections, and before the oral hearing in May 2008 at which the merits of the Commission's approach to the AEC test were hotly debated. This means that, on the basis of the AEC test, the Commission concluded that Intel's rebates made it impossible for AMD to win sales to Dell, even though in 2006 Dell had already switched a significant volume of its CPU purchases to AMD. In other words, in 2007 the Commission used a predictive test to show that a shift in demand, which had occurred in 2006, was impossible. This surprising outcome raises the fundamental question of whether there is any scope for the application of predictive tests and the consideration of potential effects in cases where the abusive conduct has already occurred and where there is evidence – the actual switching conduct of an important customer – that establishes that a customer's demand can be shifted away from the dominant firm.

It would seem reasonable to argue that, where they are known, actual effects should be given precedence over potential effects when assessing exclusionary conduct.¹⁷ In cases in which conduct is on-going and in which a regulator needs to decide whether to intervene, one can only rely on predictive tests or speculation as to likely effects. However, in cases in which the conduct has already come to an end actual effects should be discernable and measurable. In such a case, there is arguably no place for speculation or predictive tests since the outcome is already known. Indeed, such use of a predictive or speculative test could be analogous to reliance on a *per se* rule in that it can result in the condemnation of conduct without regard to its actual effects. Unfortunately, the Court of Justice never addressed this issue.

It was, however, extensively addressed by *Intel* in the Commission's administrative procedure. During the infringement period AMD had performed better than at any time in its history in terms of revenues, profitability and share growth. There was also evidence that when AMD had competitive products that outperformed Intel's, such as its chip for servers (known as Opteron), it gained market share despite substantial cash payments by Intel to OEMs who used its server chips. Conversely, there was also evidence that in key sectors AMD chips were inferior to Intel's. The Commission's answer to the argument about AMD's performance was that AMD would have done even better absent Intel's conduct – an assertion to which there is, of course, no effective response.

The Court of Justice's reliance on potential effects in a case involving actual effects is troubling for two reasons. First, *Intel* was a case in which there were very sophisticated customers competing intensely downstream and for whom the price and performance of a major input like a CPU was critical to success in the market place. It was also a case in which customers had only two potential suppliers and, thus, would presumably be reluctant to support a foreclosure strategy that –if successful– would give rise to a monopoly. Indeed, there was extensive evidence that the OEMs continuously played Intel and AMD off against each other to get better pricing, but also that quality and performance were critical factors in the ultimate choice of a chip supplier. For example, as noted above, in the case of server chips where AMD had an arguably superior product, Intel's price reductions in the form of financial contributions to the OEMs that used its products were unable to prevent AMD from greatly expanding its market share in this sector. These facts raise serious questions about Intel's ability to foreclose AMD and the impact of its discounts on customer choice. Ignoring them in favor of predictive tests or speculation about potential effects seems unjustifiable and likely to yield false positives.

Second, the reliance on potential effects in an *ex-post* case suggests that the Court has not fully rejected the *per se* approach to exclusive rebates. While the requirement to show potential effects does loosen the noose of *per se* illegality, it creates a half-way house to the extent it does not fully embrace the decisiveness of actual (as opposed to likely) outcomes in *ex-post* cases. In theory, such an approach might be justified on policy grounds if the goal were to deter even attempts at foreclosure, regardless of whether they actually succeed. However, this approach was implicitly rejected by the Court of Justice in *Intel*. For although the Court did agree with the General Court that Intel's conduct had been in execution of a strategy to foreclose AMD, it nevertheless remanded the case to the General Court so that Intel's claims about the flawed application of the AEC could be examined. If the existence of a strategy to foreclose were sufficient to infringe Article 102 the Court would not have taken this step. This is clearly positive. However, the result is that the Court of Justice has now asked the General Court to review Intel's claims about how the AEC test was applied where, in at least one case – Dell – we know that the AEC test failed to predict what actually happened.

17 In its Guidance on the Commission's enforcement priorities in Applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings (commonly known as the "Guidance Paper") issued shortly before adoption of the *Intel* decision, the Commission notes, in para 20, 6th indent, that where, conduct has been in place for a sufficient period it may be possible to rely on evidence of actual foreclosure. OJ C 45/7 24 February, 2009 at p. 10. The Commission's comment identifies the utility of actual evidence but does not seek to attribute to it greater credibility.

PATENT REFORM, INNOVATION, AND THE SCOPE OF COMPETITION POLICY



BY MARK SCHANKERMAN & FLORIAN SCHUETT¹



¹ Mark Schankerman, London School of Economics and Political Science and the Centre for Economic Policy Research. Florian Schuett, Tilburg University.

I. INTRODUCTION

Competition policy agencies have begun to pay more attention to innovation. This is a positive, and long overdue, development. Innovation is the engine of growth, and a more dynamic perspective in the formulation and application of competition policy is essential.² To date most of the attention has been on how mergers may affect the incentives for innovation of the merging, and perhaps other, parties. However, there is another important dimension – the diffusion of innovation. In the context of patented technologies, the key channels for diffusion are the licensing and sale of patents, i.e. the “market for innovation.”

While competition policy has always intervened against certain types of provisions in licensing agreements, competition authorities have recently expanded their scope of operation, with interventions against, e.g. pay-for-delay settlements and failure to comply with FRAND commitments in the context of standard-essential patents. This more interventionist approach has coincided with a perceived decline in the quality of issued patents in recent decades. Perhaps it is this decline that has led competition authorities to be more skeptical of patents in general, and to try to address the patent-quality problem by curbing patent enforcement. We believe that this approach may not be the most effective, as it attacks the symptoms rather than the causes of the problem.

The need for competition policy intervention in licensing depends in part on whether private parties, such as potential licensees or third parties, have effective means for learning about, and if necessary, challenging the validity of patents. If they do, the scope for potential anti-competitive licensing behavior by patentees may be more circumscribed. Specifically, if licensees had more confidence that patents were valid and would be upheld by the courts, the market for technology would be strengthened. Uncertainty around this issue undermines licensing and thus the diffusion process. We argue that greater certainty can be achieved by a combination of patent, and perhaps also legal reforms – which affect the quality of patent screening – and which in turn can reduce the need for expansive competition policy intervention.

The patent system is perhaps the most important institution for providing innovation incentives in modern economies. Yet, academic scholars and policymakers have begun to voice concerns that patent rights may be more of an impediment, rather than an incentive, to innovation. They point to the potential for the proliferation and dispersed ownership of patents to raise the transaction costs of doing R&D, since this may require securing many separate licenses on patented technologies owned by others. The problem is exacerbated when patents held by others are of questionable validity but it is costly and risky to use the courts to challenge them. Then, as Farrell & Shapiro (2008) argue, “weak patents” – those with low probability of being upheld as valid – may end up being unduly strong, creating greater opportunity for rent extraction.³ Public policy makers and courts have weighed in on these issues, e.g. the U.S. Supreme Court in its decision in *eBay Inc. v. MercExchange*, 547 U.S. 338, 2006 and others, and the U.S. Congress in the Leahy-Smith America Invents Act of 2011 – the most significant statutory change to the U.S. patent system in half a century.

We argue that these problems arise in large part from ineffective patent office screening, granting patents to obvious inventions that do not represent a substantial “inventive step.” If effective patent reform were introduced to address this “patent quality problem,” it is our contention that the need for competition policy intervention would be mitigated, perhaps substantially. But before public policy embarks on major reform, it is important to know whether there really is a patent quality problem, and if so, how severe it is. Not only that, but in order to design appropriate remedial reforms, we also need to identify the source(s) of the problem. Is screening by the patent office inadequate (either because the patent office is not effective or because it is applying patentability standards that are too lax)? Is it that low quality patents, which should be challenged in court, are able to pre-empt potential challenges by strategically setting licensing royalty rates? Or is the court review system ineffective because of high costs or high error rates?

In order to investigate these issues systematically, we argue that one needs an analytical framework that encompasses the full range of patent screening institutions, including the patent office, the potential for validity challenges in the presence of strategic licensing by patent owners, and the courts. Moreover, in order to be useful for assessing counterfactual policy reforms, we need to calibrate the model so it matches observed outcomes in the real world. This is what we have developed in our recent research.⁴

2 For an expression of this new direction in the U.S. Department of Justice, see various recent speeches by Makan Delrahim, Assistant Attorney General, Anti-trust Division, available at <https://www.justice.gov/atr/staff-profile/meet-assistant-attorney-general>.

3 Joseph Farrell & Carl Shapiro (2008), “How Strong are Weak Patents,” *American Economic Review*, 98: 1347-1369.

4 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2885197.

II. ILLUSTRATING THE PATENT QUALITY PROBLEM

The prevailing view is that there exists a serious patent-quality problem. One high-profile example that illustrates the problem is the Amazon “one-click” shopping patent. The year 2017 marked the expiration of Amazon’s patent on a “Method and system for placing a purchase order via a communications network” (U.S. Patent 5,960,411). This patent gave Amazon the right to exclude other online retailers from using a checkout method whereby repeat customers can complete a purchase in a single click. This conferred a considerable competitive advantage: other online retailers could only offer the same convenience if they licensed the patent from Amazon for use in their business, as Apple reportedly did for its iTunes music store.

While the one-click patent created substantial value for Amazon, many observers are skeptical as to whether it ever should have been granted. To be patentable, an invention needs to satisfy novelty and non-obviousness criteria, and the one-click patent could arguably have been struck down on either account. Yet we will never know whether a court would have considered the patent valid, as its validity was never adjudicated. Shortly after the patent was granted in 1999, Amazon sued Barnes & Noble, claiming that its “Express Lane” checkout procedure infringed the one-click patent. But the two parties wound up settling the case, so that the uncertainty about the validity of the patent remained unresolved.

Amazon’s one-click patent illustrates a number of important but troubling points. First, patents can be privately very valuable even though the underlying inventions are obvious, in the sense that the patent incentive would not be required to bring forth the investment to develop them. Second, the fact that a patent does not get litigated does not mean that it is not causing social costs: the one-click patent was used by Amazon to prevent competitors from improving their customers’ shopping experience and/or to extract licensing revenue through royalties, presumably raising the prices of final products. Third, disputes over low-quality patents tend to settle out of court, so that their legal status often remains uncertain.

Is the Amazon patent an isolated case – an egregious but unrepresentative example – or part of a more general phenomenon? How severe is the patent-quality problem? This is a surprisingly hard question. Although studies show that about half of all litigated patents for which a final court decision on validity is made are held to be invalid, there is a matter of “selection”: litigated patents are different from other patents. Patents that aren’t worth fighting over won’t be litigated; thus, litigated patents tend to be of higher than average value. But there is also reason to believe that patents litigated all the way to a court decision are of higher than average validity. As the example of the one-click patent suggests, and economic theory confirms, disputes involving low-quality patents are often settled. In that case, the court outcomes – half of all patents being judged invalid – would actually underestimate the extent of the problem, possibly by a large margin.

What is needed to measure the severity of the patent quality problem, and to evaluate counterfactual policy reforms, is an integrated model that encompasses all the stages from applications to and screening by the patent office, royalty setting and its consequences for the likelihood of validity challenges in the courts, and the outcome of such court reviews. Such a model would need to endogenize these various decisions and thus would, among other things, take account of the selection problems alluded to in the previous paragraph.

III. AN ANALYTICAL FRAMEWORK

In our recent research (see footnote 4), we developed an analytical framework that allows us to address several policy-driven research questions. First, how serious is the patent quality problem? Second, if the problem is serious, what to do about it? In other words, how can we improve patent screening? Our research is an attempt to provide the first integrated analytical framework for evaluating policy reforms in the patent area.

We focus on four key policy instruments: the intensity of patent office examination, pre-grant (application) fees paid before patent examination, post-grant (issuance/renewal) fees paid by inventions that have passed the examination, and review by the courts for patents challenged by a competitor. For most of the analysis, we assume that courts are perfect in that they invalidate patents on obvious inventions with certainty. But we also show that the key theoretical results of the model, and the simulations of counterfactual policy reforms, are similar when we allow for imperfect courts.

We are able to use this framework to conduct simulations, calibrated on U.S. patent and litigation data, in order to provide estimates of the severity of the current patent-quality problem. As we pointed out above, it is hard to know how many low-quality patent applications and grants there are. First, the quality of applications is unobservable, so the grant rate does not directly tell us how good a job the patent office is doing. Second, litigated patents are highly selected and litigation outcomes are not representative of the general population. Thus, the frequency of court invalidation does not directly tell us what share of all patents is of low quality (even if courts were perfect). By putting structure on the problem and exploiting an equilibrium framework, our approach allows us to get around both of these issues.

In our model, we impose the patentability requirement that R&D costs must exceed the profits the inventor can appropriate without a patent. This corresponds to the notion that patents should be given only to those inventions that require the patent incentive to be developed, and not to those that society would have benefited from even absent a patent (this abstracts from any social gains from the disclosure requirement in patenting).⁵ Of course, while this theoretical patentability criterion makes economic sense, the courts have struggled with the practical aspects of how to implement it. In our view, the statutory definition of patentability, and the various judicial standards for non-obviousness, novelty etc., reflect an attempt by the courts to do this, the basic presumption being that if an invention is obvious to those skilled in the relevant scientific arts then it is probably cheap to develop, and one does not need a patent to induce it.

In our framework, inventors hold private information about the quality (validity) of their patents. They know more (even if not perfectly) about whether an invention was obvious than their competitors and third parties, such as the patent office and the courts – whether the invention was easy to make, or whether it took years of R&D to come up with. After R&D investment, an inventor chooses whether to pay a pre-grant fee and, if subsequently approved (patent office screening is imperfect), whether to pay a post-grant fee to activate the patent (issuance and renewal fees). If the patent is activated, the inventor may choose to license the invention to the competitor, and the competitor chooses whether to challenge the validity of the patent in court. The baseline model has a perfect court that always invalidates an obvious patent and upholds a non-obvious one. In an appendix and in the simulations we also analyze a generalized version of the model in which both the patent office and the court make two-sided errors – granting/upholding an obvious patent and rejecting/invalidating a non-obvious one. Formally, our model is a signaling game in which each decision by the inventor can reveal information about the invention type, and the competitor Bayesian updates.

Our model yields three key theoretical, but policy-relevant, results. First, the equilibrium always involves a degree of randomization in patentee behavior – in particular, inventors with non-obvious inventions (“high type”) always apply for patents and charge high license fees, while inventors with obvious inventions that should not be patented (“low types”) either randomize between offering low and high license fees, or between applying and not applying for patents. Second, *even if the courts are mistake-free*, they cannot eliminate all bad patents. This is because of selection into litigation: in equilibrium not all low-type patents are challenged by the competitor. This result raises serious doubts about over-reliance on the court system to weed out bad patents.

Third, we show that the socially optimal structure of fees involves frontloading, i.e. relying on pre-grant rather than post-grant fees. The intuition is that the low type prefers post-grant fees to pre-grant fees more strongly than the high type because the low type has a smaller chance of passing examination. As we explain later, this is very different from the way in which patent fees are now structured.

Equipped with this theoretical framework, we calibrate the model so that the equilibrium predictions of the model match observed outcomes in the U.S. from the patent office and litigation in the courts. We simulate the calibrated model in order to assess the severity of the current problem with patent quality and screening, and to study the welfare impact of various counterfactual policy reforms. The simulations reveal some striking findings about the patent quality problem: 75 percent or more of patent applications are “low quality” in the economically relevant sense that they are made on inventions that would be developed even without patent rights (as a society, we don’t want to grant patents unless they are needed to induce the innovations). Not only that, but the patent office screens out about 30 percent of these low-quality applications. As a consequence, we conclude that between 65 and 81 percent of granted patents are invalid. These findings highlight the crisis in patent screening and the need to develop effective policies to address it. Indeed, many of the problems associated with the patent system – such as patent thickets and patent assertion entities, pejoratively referred to as “patent trolls” – can be traced back to a lack of patent quality.

IV. HOW TO FIX IT?

How do we improve patent screening? Which policy reforms could help ensure that deserving inventions receive patent protection and undeserving ones don’t? Our theoretical framework emphasizes the importance of integrated thinking which captures the linkages between the different parts of the system: patent office screening, licensing, and validity challenges in the courts. In recent years, reform efforts have focused mainly on the courts. For example, the Leahy-Smith America Invents Act (“AIA”) of 2011 and recent Supreme Court decisions have implemented a variety of measures to make it harder to enforce low-quality patents.

We think that this focus on the courts to limit the patent quality problem is misplaced, and that renewed emphasis should be put on patenting fees and patent office examination. The reason is that screening of patent validity by the courts must be initiated by someone. Usually, the

⁵ Importantly, this perspective is in line with the rationale courts and legal scholars typically give for the non-obviousness requirement in patent law (Rebecca Eisenberg (2004), “Obvious to Whom: Evaluating Inventions from the Perspective of PHOSITA,” *Berkeley Technology Law Journal*, 19(3): 885-906). For example, in the landmark case of *Graham v. John Deere*, (383 U.S. 1 (1965)) the U.S. Supreme Court stated: “The inherent problem was to develop some means of weeding out those inventions which would not be disclosed or devised but for the inducement of a patent.”

only one in position to do so is the (alleged) infringer, who counters an infringement suit by challenging the validity of the asserted patent. Often the infringer has no incentive to challenge at all: given the enormous costs of patent litigation, only high-value patents will be worth fighting over. That low-value patents will usually not be challenged does not mean they don't cause damage, however: our research suggests that only the 10 percent most valuable patents are potentially worth challenging, due to the costs of litigation. Would-be infringers are better off taking a license, resulting in higher prices for consumers.

Even when a patent is sufficiently valuable, the alleged infringer's incentives to challenge typically aren't aligned with society's interests. First, the holder of a high-value, but low-quality, patent sometimes simply pre-empts challenges. This can be achieved by offering a license fee that leaves the infringer better off accepting the patent license rather than incurring the cost of going to court, even if the infringer is reasonably sure of winning at trial. (Such a strategy is said to be followed by many patent assertion entities.) But it is precisely in such a case, where the patentee reveals himself as having a low-quality patent by charging this "challenge pre-empting licensing fee," that society would want a challenge to be brought.

Second, even if this challenge pre-emption licensing strategy is not adopted by the patent holder, the infringer is uncertain about the prospects of winning at trial and – from society's point of view – may challenge either too little or too much. The conventional wisdom is that there are too few patent challenges, due to the "free-rider" problem: when there are many potential licensees, all of them would benefit from the patent being invalidated but each of them would prefer somebody else bear the cost of the challenge. This is the view emphasized by Joe Farrell & Carl Shapiro, among other leading antitrust scholars, and it has gained wide acceptance.

Yet our research shows that this view is too narrow, and that the market can also generate socially excessive patent challenges. The argument is easiest to see in the case where there is a single potential licensee, so that free-riding concerns are absent. The incentive to challenge (whether private or social) is determined by a comparison between the gains from the patent being invalidated and the costs of the challenge. For the licensee, the gain consists in eliminating the competitive advantage the patent confers on its holder; the cost consists in its own litigation expenses. For society, the gain consists in eliminating distortion due to higher prices from patents (deadweight loss); the cost consists of the sum of litigation expenses for both parties – the licensee and the patent holder. Economics suggests that the competitive advantage typically exceeds the deadweight loss. Thus, in this case the private incentive to challenge exceeds the social incentive.

The ambiguous nature of the comparison between social and private incentives to challenge makes it hard to design appropriate policies targeting litigation. At its core, the problem stems from the fact that challenges must be initiated by a potential licensee, who acts strategically, and whose incentives in general aren't aligned with society's interests. By contrast, patent office examination has the advantage of screening all applications. Low quality patents that are screened out by the patent office do not impose costs on society. They lead to neither royalties that inflate price, nor costly litigation. In our counterfactual policy experiments using the simulated model, we find that ramping up the rigor of examination from 30 percent to 50 percent would generate a sizeable increase in the welfare benefits from innovation. In our view these are strong arguments for focusing reform efforts on improving patent examination.

The counter argument is that improving examination would be expensive. The USPTO received almost 600,000 patent applications in 2015. Tighter examination would mean spending more time on each of them, which would require hiring more examiners. However, there is a way to finance tighter examination that wouldn't cost taxpayers a cent, and also wouldn't undermine high-quality applicants. Currently patent office fees are relatively low and largely back-loaded: the bulk of them is due after the patent is granted. A typical U.S. patent costs the applicant \$1,740 in pre-grant fees and – if renewed to full term – about \$13,560 in post-grant fees. Decreasing post-grant fees while simultaneously increasing pre-grant fees would disproportionately hit low-quality applicants. This is because low-quality applicants have a lower probability of passing examination; thus, many of them never pay the post-grant fees. Frontloading fees – shifting fees forward from post-grant to pre-grant – would thus raise more money from low-quality applicants. At the same time, it could be done in such a way as to keep the expected total fee payments for high-quality applicants constant.

This argument of course relies on low-quality applicants continuing to file applications, in spite of higher expected fee payments. But in some sense, this is a worst-case scenario: if low-quality applicants instead drop out, even better – we have made screening more effective without spending more on examination!

What is more, improving examination may not have to involve hiring lots of new examiners. Some of it could be achieved through better databases and improvements in the patent office's IT systems. Economics also suggests that improved performance could be achieved by putting in place better incentives for patent examiners, e.g. by tying bonuses to the quality of examination (not just quantity, as is the case now) through random review of their decisions. Improvement may also need to involve developing stricter patentability standards, which would require actions by the Congress and/or the courts.

The table below summarizes the simulated welfare effects of three counterfactual policy reforms: (1) frontloading fees and using the extra fee revenue generated to intensify patent examination (so the reform is revenue-neutral); (2) the introduction of a post-grant review within the patent office, such as the Patent Trials and Appeals Board (“PTAB”) instituted by the AIA; and (3) moving to a registration system without any substantive patent examination.

The first and second columns show the results under the assumption that courts are perfect. Frontloading would lead to a substantial welfare gain of close to 2 percent. Post-grant review, which involves replacing courts with a substantially cheaper, but also less accurate challenge procedure, generates even greater welfare gains. By contrast, a registration system would be associated with major welfare losses. The table also shows the percentage of invalid patent holders that pre-empt challenges by charging low royalties. Importantly, there is no correlation between such “trolling” and the size (or sign) of welfare changes. Thus, focusing reform efforts on reducing trolling would seem to be misguided.

The third and fourth columns show the results under the assumption that courts are only slightly better than the patent office at distinguishing valid from invalid patents. The results are qualitatively similar in this more realistic scenario. The effects of all patent reforms considered are actually even larger in size than under perfect courts. Of course, one can envision other ways to model the post-grant review under the PTAB – e.g. the quality of the review could be changed to be more like the courts (here it is modelled as equivalent to the initial patent office review). But the analytical framework we develop can easily be used to evaluate the impacts of such alternatives.

Policy Reform	Simulations of Counterfactual Policy Reforms			
	Perfect Courts		Low quality courts	
	% Welfare Gain	Prob. of “trolling”	% Welfare Gain	Prob. of “trolling”
Status quo	---	0.85	---	0.65
Frontload fees	1.90	0.76	3.89	0.34
Post-grant review	3.66	0.9	6.24	0.92
Pure registration system	-3.10	0.86	-4.2	0.77

Note: Welfare change is relative to the case of no invention. Prob. of trolling is the probability that a low type (invalid patent) charges a royalty rate to pre-empt a validity challenge.

V. CONCLUDING REMARKS

Our paper develops a framework to examine how governments can improve the quality of patent screening, incorporating four policy instruments: patent office examination, pre- and post-grant fees, and challenges in the courts. Perhaps the most important message to emerge from the theoretical analysis is that patent office fees should be frontloaded, and that we should not overly rely on the court system to screen patents *ex-post*. Even if courts are mistake-free, they cannot eliminate all bad patents that are issued because, in equilibrium, not all such patents are challenged. The calibrated simulations of the model indicate that the patent quality problem is real: the vast majority of patent applications are on inventions that would be developed even without patent protection, and only about 30 percent of these are effectively screened out by the patent office.

But all is not lost. Our simulations of counterfactual policy reforms highlight some feasible changes that will improve welfare: most importantly, intensifying patent office examination, front-loading patent fees, and introducing post-grant review (or reducing litigation costs by other means).

Finally, we believe that the framework developed in this paper can also be used to study the welfare impacts of legal reforms that may help in improving patent quality and screening, such as adopting the “loser pays rule” for litigation costs, changing the presumption of validity used by courts, and patent litigation insurance, all of which have featured prominently in the public debate.



WHAT IS THE SCOPE FOR CHOICE AND COMPETITION IN EDUCATION?

BY PROFESSOR ALLAN FELS & DR. DARRYL BIGGAR¹



¹ Allan Fels, Professorial Fellow, University of Melbourne. Dr. Darryl Biggar, Special Economic Advisor of the Australian Competition and Consumer Commission and the Australian Energy Regulator.

I. INTRODUCTION

Is there scope for greater reliance on competitive market forces in the provision of education services?² We argue that whilst there appears to be much scope for choice and competition in education, and certainly greater scope than exists at present in many countries, there are also some significant, often ignored, limitations on the scope for choice and competition if a major goal of education policy is to promote “equality of opportunity.”

II. CHOICE AND COMPETITION IN THE EDUCATION SECTOR

Most governments are heavily involved in funding education services, especially in primary and secondary education. In addition, in most countries the government is not merely the primary funder of education, but the primary *provider* of education services. Furthermore, in many countries competition is not actively encouraged. In fact, in many cases, competition between government education providers is often deliberately and strictly limited, particularly by granting each education provider a limited geographic monopoly over the provision of education services to children within each school’s zone.

Why? Is there something about the provision of education services which *requires* a restriction on competition? Is conventional in-the-market competition somehow in opposition to the objectives the government is trying to achieve?

To answer this question we need to look at the underlying economic problem (or “market failure”) which gives rise to a need for government intervention in the first place. As we will see, that underlying economic problem gives rise to a motivation for substantial government subsidies for the education sector. We will proceed by exploring the scope for reliance on choice and competition in education in the context of government subsidies

A. Why are Governments Involved in Education?

Three common rationales are offered for government involvement in education:

1. There are economic and political **spillover effects**:

- A tax dividend effect – if education results in higher income in the future, individuals are not able to capture all of the benefits due to the effect of taxes. A government subsidy upfront may offset the distorting impact of taxes;
- A productivity spillover – according to this theory, education increases the productivity of the individual, but also increases the productivity of individuals with whom that person interacts;
- A democratic spillover – having an educated citizenry allows the population to understand and participate in public decision-making and democratic processes, which leads to better democratic outcomes overall. As Stiglitz (1988) observes: “A society in which everyone can read can function far more smoothly in which few can read.”³

Stiglitz (1988) expresses some skepticism about the magnitude of these spillover effects especially compared to the size of the benefit the individual receives from education:

² This paper draws on two wider papers by Professor Allan Fels & Dr. Darryl Biggar, “The Role of Choice and Competition in Education,” CRESSE (2018) Conference, Greece and “The Role of Choice and Competition in Public Markets,” OECD Global Forum on Competition, December 7-8, 2018. See: [https://one.oecd.org/document/DAF/COMP/GF\(2017\)7/en/pdf](https://one.oecd.org/document/DAF/COMP/GF(2017)7/en/pdf).

³ Stiglitz, Joseph E., *Economics of the Public Sector*, 2nd edition, Norton & Company Inc, New York, 1988. For more extensive references see Fels & Biggar, opus cit, 2017. In this paper we rely heavily on Stiglitz.

There is a large private return to being able to read, and even in the absence of government support, almost all individuals would learn this and other basic skills. Indeed, most individuals would go far beyond that. The question is: given the level of education that individuals would privately choose to undertake were there no government subsidy, would further increases in education generate any significant externalities? There is no agreement concerning the answer, but the case for government support based on these kinds of externalities seems, at best, unproved.

2. **Capital market imperfections.** Although it is theoretically possible for citizens to borrow to invest in education against their future lifetime income, in practice (due to moral hazard and asymmetric information) such contracts are difficult to enforce. As Stiglitz (1988) notes: "Private lenders are not, for the most part, willing to lend to finance education, and hence those without funds of their own (or their parents') would be denied access to higher education without some assistance from the government."

Governments may have a comparative advantage in using the tax system to compel students to repay loans. But this is an argument for government provision of educational loans as opposed to subsidies. Many countries have a system of government-backed lending for education services, particularly for higher (tertiary) education. But such systems typically operate alongside, instead of in place of, direct government funding and provision of primary and secondary education services.

3. The desire to promote **equality of opportunity.**

Stiglitz explains this last concept as follows:

The primary justification for public support of education arises from concern about the distributional implications of the private financing of education. Richer individuals will want to spend more on the education of their young, just as they spend more on cars, homes, and clothes. There is a widespread belief that the life-chances of a child should not depend on the wealth of his parents or the happenstance of the community in which his parents live.

We note in passing that promoting equality-of-opportunity may be more than a fairness or equity issue; it may also promote economic growth but we will not pursue this issue here.

For the purposes of this paper we will assume that a primary motivation for government involvement in education is to promote a degree of equality of opportunity. This motivation gives rise to government subsidization of at least the core primary and secondary education services – to ensure that all citizens have access to a certain quantity and quality of education services regardless of their background or income.

B. Is Competition between Educational Institutions even Feasible?

In principle, if it were possible to harness the forces of competition in education without compromising the objectives above, education might be delivered like many other services. We would expect a wide range of competing educational institutions meeting the needs of a wide range of students, and innovating to develop new educational techniques, making efficient use of new technology and other inputs, responding to the changing desires of parents and the needs of the broader community, and doing so with the minimal claim on government resources. But to what extent is this vision in conflict with the objectives identified above? To what extent can we rely on choice and competition to deliver desirable outcomes in the education sector?

First let's address the question whether effective competition between education service providers is even feasible. In principle, an assessment of the scope for competition in education services requires a formal competition analysis – that is, an analysis of the product and geographic dimensions of the market, barriers to entry and exit, and other factors such as transparency of prices and quality, vertical integration, and barriers to customer choice. However, to keep this discussions brief we merely make the following observations.

There do not appear to be inherent structural barriers to achieving effective competition between educational institutions in most large towns and cities. In particular, there do not seem to be overwhelming economies of scale or scope in the provision of traditional primary and secondary education within a reasonable geographic market. Most urban areas are able to sustain many traditional educational institutions within any given 5-10 kilometer radius. But even this radius is too far to walk; if competition is to be effective, consideration may need to be given to how students who make a choice of school will be able to travel to that school, perhaps by ensuring access to public transport.

The range of sizes of different schools which we can observe in operation suggests that relatively small schools (with less than, say, 200 students) are able to co-exist and compete with quite large schools (with more than, say, 1000 students). It appears, therefore, that, with the exception of relatively small towns, most urban areas should be able to sustain active competition in education services in most areas, even without students having to travel very long distances.

In order for choice and competition to be effective in improving overall outcomes, schools must have the incentive and ability to respond to market conditions, and to adapt and innovate in response to financial incentives. Historically, in countries where the education sector has been dominated by government providers, the scope for individual schools to exercise autonomy, independence and innovation has often been strictly limited. Government schools have historically not had the incentive nor the ability to respond to market conditions.

In other sectors, of course, private, for-profit firms typically have strong incentives to adapt and innovate where doing so will increase the overall stream of profits. We do not see any particular reason why private for-profit firms would not have the same incentive in the education sector. Across the OECD, of the 12 percent of students who are enrolled in private government-dependent schools, around 8 percent are educated in for-profit schools.

But for-profit schools remain the minority and in many OECD countries do not exist at all. Across the OECD, the majority of students enrolled in private government-dependent schools attend not-for-profit institutions – either religious not-for-profit schools or other non-religious, but not-for-profit schools. There may be good reasons why not-for-profit institutions survive and thrive in the provision of public services. For example, where quality is difficult for customers to measure, a commitment to a not-for-profit structure may provide an assurance and implicit commitment to customers that quality will not be cut *ex-post*.

The effectiveness of both for-profit and not-for-profit structures depends on the quality of the governance and oversight. High-quality governance is typically easier to ensure in the case of a for-profit structure where the owners have a direct financial incentive in the performance. But some of these incentives for high-quality governance could, in principle, be reproduced where the members of the Board of the school have a direct personal interest in its success, such as parents who have children at the school.

A full discussion of the merits of for-profit and not-for-profit firm structures in the education sector is beyond the scope of this paper. We merely note that, whatever structure is chosen, for competition to be effective, education service providers must be willing and able to adapt and respond to market conditions, to meet the needs of customers, and to compete to attract students. Both for-profit and not-for profit providers seem capable of meeting this condition.

For choice and competition to be effective it is essential that schools have the incentive and ability to adapt. In addition, underperforming schools must be allowed – and indeed must be required – to exit the market. No matter how profitable and innovative their product offering, new entrant education providers will be deterred from entering the market if incumbents are able to persistently earn a below-market cost of capital. Failing schools must be allowed to exit the market. There is a question as to how this applies to not-for-profit schools. Not-for-profit institutions (which are typically not subject to the threat of takeover) may be immune from the normal capital market disciplines. Among other things, not-for-profit institutions may be able to persistently earn below-market return on invested capital. This could in principle act as a barrier to entry. These questions remain to be explored.

Of course, competition does not deliver desirable results unless the customers are willing and able to make choices in their own or their children's best interests. In our view, the majority of parents are willing and able to make an effective choice of schooling on behalf of their children. However, some parents do not have the capacity or the desire to do so and, in these cases, special support is required. In particular, parents with limited education or ability to make an effective choice may require assistance or support in the choice of school for their children. Otherwise, there is a risk that disadvantage may become entrenched from one generation to the next, undermining equal opportunity objectives.

It is likely that consumers of education services will face something of an information problem. Consumers are likely to have difficulty in observing the quality of the education provided – even for those consumers who have attended a particular school, it can be hard to judge the relative quality of education (i.e. the counterfactual – how would the outcomes be different if I had attended a different school?). In any case, the quality of the education may be a personal or idiosyncratic experience (would the outcome be different if I attended the same school but was assigned different teachers, or met different classmates?). It can be even harder assessing the quality of education at a school without any direct personal experience.

The results obtained by students in standardized testing are relatively easily obtained. But such measures are only, at best, imperfect measures of the quality of education, since such results depend in part on the *ex-ante* or intrinsic ability of the students and not solely on the quality of the teaching.

This information asymmetry problem gives rise to the need for independent objective review, assessment, and accreditation services. These services could perhaps be provided by the market or, alternatively, could be provided by the government.

In 2006, the British government concluded that for choice-based systems to be successful in raising educational standards for all, the following conditions were required:

- Parents – particularly, parents of disadvantaged children – need to be provided with high quality information, guidance and advice about the options open to them;
- Help with the costs of making choices, such as covering the costs of transport to alternative schools, should be provided by those who need it; ... and
- Above all, early effective action should be taken to tackle failing or poorly performing schools and increase the supply of good schools. Ultimately, the greatest safeguard against adverse effects on social inequalities and segregation is to make sure there are more good schools ... Evidence ... suggests that where schools face no financial implications from not attracting pupils there is only a limited behavioural response to choice.⁴

In summary, achieving effective competition between schools may require specific government action, such as the provision of comparator information on school quality, provision of transportation services for students so that they can travel to the school of their choice, and mechanisms to ensure that schools have the ability to respond to financial incentives, and action to ensure that underperforming schools are able to (and indeed required to) exit the market. Provided that these issues are addressed there do not seem to arise insurmountable barriers to competition between schools. But can this choice and competition be made compatible with government subsidies and the government objectives for education set out above?

C. Making Competition Compatible with Subsidization of Education

The scope for choice and competition depends on how the government subsidies are paid.

Many countries have considered, trialed, or adopted **school voucher** schemes under which customers have a choice of school and the government educational subsidy is paid directly to the educational institution chosen by the end-user.

School voucher schemes are an example of a fee-for-service subsidy payment. Provided the revenue associated with the voucher exceeds the marginal cost of taking an additional child, the voucher scheme, coupled with choice, gives rise to a high-powered incentive on schools to attract students and therefore to achieve high levels of productivity, innovation, and responsiveness to the needs of customers.

But, there are also problems with the fixed-fee-per-service model.

III. THE TOPPING-UP PROBLEM

It is widely believed that access to education is a fundamental component of promoting equality of opportunity. It follows that education services should not be allocated on the basis of parental income or wealth. As noted earlier, Stiglitz comments that “the life-chances of a child should not depend on the wealth of his parents or the happenstance of the community in which his parents live.”

However, where choice and competition between schools is permitted, it is highly likely that some schools will be in demand. These schools may choose to, or be forced to, limit enrolments. The natural mechanism to balance supply and demand in this context is to increase

⁴ UK Government (2006), page 5. Another condition was a prohibition on cream-skimming: “Schools should be prevented from ‘cream skimming’ ... by putting in place funding regimes that reflect the higher costs of teaching some groups and by using regulation and statutory guidance to prevent unfair selection.” This issue is discussed further below.

the price – in this case, the tuition fee. This implies charging an additional fee to the end-customer over-and-above the government subsidy. We refer to this as “topping up.” In fact, under a system of unrestricted competition between schools, many schools will seek to, or be forced to, ration entry through the price system. Some schools will be considerably more expensive than others. These may be the schools which provide the greatest life-opportunities.

In this world, the outcome of competition is likely to be – at least to some extent – an allocation of places at the schools which offer the most opportunity to those with the greatest ability to pay. In fact, for schools which choose to limit enrolment, the supply curve is, in effect inelastic. An increase in the government subsidy would be expected to result in an equivalent dollar-for-dollar increase in the price for education. In the absence of further intervention, the provision of a government subsidy coupled with the right to choose the educational institution undermines the objective of equality of opportunity.

Can we retain the benefits of competition while preserving the objective of equality of opportunity?

Many countries which have chosen to implement a voucher system for school choice (such as Chile or Sweden) have explicitly sought to prohibit “top up” funding.

But prohibitions on parental contributions are not easy to enforce. Schools which seek to ration entry can request a range of supplementary inputs that parents are expected to provide. The simplest approach is to ask parents to make a “voluntary” contribution to the school. Parents can also be expected to volunteer their time to the school. Although not as obviously discriminatory, such rules can have a discriminatory effect where low-income families have less leisure time to devote to volunteering at the school and where well-educated parents can offer more highly-valued inputs (such as tutoring students).

If rationing through price is explicitly prohibited, how are schools to balance supply and demand? One possible approach is to ration through waiting lists or through lottery. In our view, neither of these approaches is satisfactory. In some countries over-subscribed schools are allowed to ration through waiting lists. However, in one study it was found that the children who attended the school were overwhelmingly born in the first half of the year since children could be entered on the waiting list as soon as they were born. It does not seem satisfactory to ration access to education (and life-opportunities) by the happenstance of birthdate.

Some countries with school voucher systems (such as Chile) explicitly prohibit rationing by schools which participate in the voucher program. Schools which participate in the voucher program are required to accept all applicants. This is a fairly drastic intervention, but in our view is worth considering.⁵

To summarize, if equality-of-opportunity goals are to be achieved, access to educational opportunities should not be allocated on the basis of parental income or wealth. Although we have no objection to some, modest level of school fees, the ability of schools to ration access by charging above a minimum threshold would be expected to result in allocation of educational opportunities on the basis of parental income and therefore must be restricted. This will likely require active enforcement. We do not support allowing rationing by queuing or by lottery. Instead, we prefer a mechanism in which all participating schools are required to accept all applicants (perhaps with some notice period, to mitigate problems from large swings in numbers).

IV. THE CREAM-SKIMMING PROBLEM

The other primary concern with encouraging school competition using vouchers is that of cream-skimming: the incentive for schools to identify and deny entry to high-cost-to-educate students.

As we noted earlier, potential students differ widely in how much it will cost for the school to achieve its desired objectives with that student. As long as the voucher is for a fixed fee, independent of the characteristics of the student, each school has a strong incentive to compete vigorously for students which can achieve their desired objectives at the lowest possible cost; and to actively resist taking students for which achieving the desired objectives will be very high-cost.

⁵ The primary argument against an obligation to take all applicants is that doing so would lead to a dilution of quality. But why is this the case? Do the key inputs necessary for the provision of a high-quality education exhibit decreasing returns to scale? Presumably good teachers can be trained and good management and governance practices extended over multiple schools. If there are diminishing returns to scale in school size, then new schools can be created. Indeed, it is plausible that there are economies of scale in education, with larger institutions able to deliver higher-quality services more efficiently. A requirement to accept all applicants would eliminate any premium for exclusivity but exclusivity is unrelated to quality of education. On balance we consider an obligation to take all applicants worth considering.

But what are the objectives that each school wishes to achieve? In a world of school choice and competition, the objectives of each school presumably reflect the desires of customers. What do prospective customers (students or their parents) look for in a school?

Where schools are assessed primarily on the basis of the academic performance of graduating students on standardized exams, the cheapest way for a school to achieve good outcomes is to select entering students with high innate ability. This results in quite strong competition for the very best students. Students of average ability may find the range of schools willing to accept them is limited, while students of low ability may find no opportunities at all. This system tends to favor parents who have well prepared their children, perhaps through tutoring, at an early age. It therefore seems to conflict with the objective of equality of opportunity.

How can this problem of cream-skimming be addressed? One possible approach is to carefully differentiate students into different groups *ex-ante*, and to match the funding revenue to the likely cost of educating each student. This has its own challenges and takes us beyond the scope of this brief discussion.

V. ENHANCING CHOICE AND COMPETITION IN THE EDUCATION SECTOR

In short: we believe that effective competition in education services is possible, despite the presence of government subsidies and the need to pursue the basic government objectives identified above. However, perhaps unsurprisingly, specific and relatively strong controls on that competition are essential. Specifically, in our view, schools which receive government subsidies must not be allowed to compete on price, and probably should not be allowed to deny entry to any applicant. These are relatively severe interventions. In addition, the government must pay careful attention to differences in costs of educating different groups of students. Such differentiation may be tricky. Although we consider that this is not inevitable, there remains a risk that promoting competition between schools will enhance social segregation and undermine, rather than promote, social cohesion. We remain hopeful that these problems can be overcome.

At the same time there are large potential benefits. In our view there remains significant scope for school choice to significantly enhance the incentives on schools to provide educational services as efficiently as possible, and to innovate, to provide the variety of services that parents and students desire, and to respond to changing market demands. Although the flame of competition between schools must be tightly circumscribed, it is by no means necessary to extinguish it entirely.



PUBLIC AND PRIVATE ANTITRUST ENFORCEMENT FOR CARTELS: SHOULD THERE BE A COMMON APPROACH TO SANCTIONING BASED ON THE OVERCHARGE RATE?¹

BY YANNIS KATSOULACOS², EVGENIA MOTCHENKOVA³ & DAVID ULPH⁴



¹ We are grateful to Joe Harrington, Frederic Jenny, Tom Ross, Maarten Pieter Schinkel, Simon Roberts as well as the participants of the 12th Annual CRESSE Conference (July 2017) and the GDEC-CRESSE International Workshop on Advances in Competition Policy (Rio de Janeiro, November 2017), for helpful comments on our papers related to the subject of antitrust sanctioning. Also, we are grateful to the Tinbergen Institute, Vrije Universiteit Amsterdam, Short-term Visitor Program for financial support.

² Department of Economic Science, Athens University of Economics and Business, Patission 76, Athens 104 34, Greece. Email: ysk@hol.gr.

³ Department of Economics, Vrije Universiteit Amsterdam, TILEC and Tinbergen Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: emotchenkova@feweb.vu.nl.

⁴ School of Economics and Finance, University of St Andrews, KY16 9AR, Scotland. Email: du1@st-andrews.ac.uk.

I. INTRODUCTION

The imposition of sanctions has been regarded as the most important *ex-ante* public enforcement instrument that Competition Authorities (hereafter “CAs”) can use in response to antitrust and, more specifically, cartel violations.⁵ It is complemented by private enforcement in the form of private damage actions. In principle, by imposing sanctions for infringements, public enforcement’s main objective must be to deter violations (*deterrence effect*) and to induce non-deterred colluding firms to charge lower prices (*price effect*), while private damages focus on compensating those who have suffered harm. Clearly, each method can contribute to the objectives of the other. Public enforcement can facilitate and stimulate private damage actions and private damage actions can contribute to deterrence and provide incentives for customers to discover and report price-fixing. This paper reviews the recent literature pointing to the ineffectiveness, in terms of their welfare impact,⁶ of monetary penalty schemes currently used by CAs⁷ and argues the case for CAs switching to a more effective penalty regime in which the *penalty base* continues to be the currently dominant penalty base of cartel revenue but where, in contrast to current practice, the *penalty rate* is based on the cartel overcharge – which is often estimated in order to calculate damages in private damage actions.

The extensive and still growing literature by economists on monetary penalty regimes⁸ examines and contrasts alternative types of such regimes, concentrating on a comparison of their welfare properties. It is nevertheless recognized that, while this comparison is very important, in order for it to have practical policy significance a number of other policy-relevant dimensions have to be assessed and compared. Specifically, a more complete comparison must take into account the following three dimensions/assessment criteria:

5 The theory of sanctioning on dominant firm abuses is still undeveloped. Other important *ex-ante* instruments of competition law enforcement in the area of cartels are the prohibition of *facilitating practices* (which can increase the viability of cartels) and the use of merger policy (to reduce the likelihood of cartels emerging after mergers). *Ex-post* measures include the improvement of detection and prosecution rates, the adoption of measures to prevent recidivism and the application of leniency policies.

6 The continuing high prevalence of cartels across markets, confirmed by extensive empirical evidence also testifies to this. See e.g. Levenstein & Suslow (2011), “Breaking Up Is Hard to Do: Determinants of Cartel Duration,” *Journal of Law and Economics* 54, 455-92; Levenstein & Suslow (2012), “Cartels and Collusion - Empirical Evidence,” Ross School of Business Paper No. 1182, available at <http://ssrn.com/abstract=2182565>; Levenstein & Suslow (2014) “Price fixing hits home: an empirical study of price fixing conspiracies in the US,” mimeo, University of Michigan, Jan. 2014; Schinkel M.P. (2007), “Effective Cartel Enforcement in Europe”, *World Competition* 30(4), 539-572; Veljanovski, C. (2007), “Cartel Fines in Europe: Law, Practice and Deterrence”, *World Competition* 29; Connor & Lande (2008), “Cartel Overcharges and Optimal Cartel Fines,” in S.Waller (ed), *Issues in Competition Law and Policy*, Vol 3, AMA Section of Antitrust Law, Chapter 88; Allain et al. (2011), “The determination of optimal fines in cartel cases: Theory and practice,” *Concurrences - Competition Law Journal* 4-2011, 32-40; Boyer & Kotchoni (2015), “How Much Do Cartel Overcharge?,” *Review of Industrial Organization*, 47, 119-153; or Spagnolo G. and C. Marvão (2016), “Cartels and Leniency: Taking stock of what we learnt”, in *Handbook of Game Theory and Industrial Organization*, by L. C. Corchón, and M. A. Marini (Eds.), Edward Elgar Publishing, 2016, for an overview. This, of course, is not the only potential inadequacy in public enforcement and recent literature has also pointed to the ineffectiveness of monetary penalties, as currently applied, in inducing desirable price effects – see Bageri et al. (2013), “The Distortive Effects of Antitrust Fines Based on Revenue,” *The Economic Journal*, 123 (572), 545-557; Katsoulacos & Ulph (2013), “Antitrust Penalties and the Implications of Empirical Evidence on Cartel Overcharges,” *The Economic Journal*, 123 (572), 558-581; Katsoulacos et al (2015), “Penalizing Cartels: The Case for Basing Penalties on Price Overcharge,” *International Journal of Industrial Organization*, 42, pages 70-80.

7 There is a variety of different types of sanctions with different emphasis placed on each type over time and in different countries. Here we concentrate on monetary penalties on corporations. The other main types of sanctions in public enforcement are: financial penalties on managers involved in price-fixing, criminal sanctions/imprisonment of individuals involved in price-fixing, debarment of individuals involved in price-fixing, from further employment in a position from which they could again violate antitrust laws. See for a review Katsoulacos et al. (2017), “Penalizing on the basis of the severity of the offence: a sophisticated revenue-based policy for sanctioning cartels,” Tinbergen Institute Discussion Paper Series; vol. 17, no. 120/VII.

8 See e.g. Harrington (2004), “Cartel Pricing Dynamics in the Presence of an Antitrust Authority,” *The Rand Journal of Economics* 35, 651-673; Harrington (2005), “Optimal Cartel Pricing in the Presence of an Antitrust Authority,” *International Economic Review* 46, 145-170; Buccirosi & Spagnolo (2007), “Optimal Fines in the Era of Whistle blowers - Should Price Fixers Still Go to Prison?,” in *The Political Economy of Antitrust*, by V. Goshal and J. Stennek (Eds.), Elsevier: Amsterdam; Harrington (2010), “Comment on Antitrust Sanctions”, *Competition Policy International*, 6, 41-51; Houba et al. (2010), “Antitrust enforcement with price-dependent fines and detection probabilities,” *Economics Bulletin*, 30(3), 2017-2027; Bageri et al. (2013), *supra* note 6; Katsoulacos & Ulph (2013), *supra* note 6; Dargaud et al. (2015), “Cartel deterrence and distortive effects of fines,” *Journal of Competition Law and Economics*; Katsoulacos et al. (2015), *supra* note 6; Katsoulacos et al. (2018), “Sophisticated revenue-based cartel penalties vs overcharge-based penalties,” Amsterdam: School of Business and Economics, Research Memorandum; vol. 2018-1 2018); Bos et al. (2018), “Does enforcement deter cartels? A tale of two tails,” *International Journal of Industrial Organization*, 59, 372-405.

- (i) **Implementability.** This involves considerations relating to the administrative cost of the penalty regime,⁹ the extent to which it minimizes delays in the CA enforcement process¹⁰ and the extent to which it minimizes the costs of appeals in the judicial review process.¹¹ The latter will be higher the more the appeals are induced against the CA's penalty decisions by a penalty regime. The number of appeals will be greater the more likely it is that the penalty regime can lead to estimation errors and/or when penalty decisions can be easily challenged as discriminatory.
- (ii) **Transparency/Certainty.** Penalty regimes differ in terms of how easily and accurately firms can predict the fine they will be facing IF they are prosecuted and are found to have violated the law. When firms cannot predict or estimate the penalties that the CA will set were it to investigate and condemn their conduct, this represents a level of uncertainty and lack of transparency.¹² We consider transparency/certainty to be a desirable feature in a penalty regime, taking the position of a large number of jurisdictions (including the EC, U.S., Canada and Brazil), that to reach deterrence targets agencies must rely on the threat of severe penalties coupled with a significant fear of detection. While it is known that in a few cases agencies have adopted the view that some uncertainty can improve deterrence, when detection rates are low and the severity of penalties is constrained, this approach is recognized as having serious downsides.¹³
- (iii) **Welfare properties.** As noted above, it is on these properties that the economic literature has concentrated. While the traditional literature identified first-best optimal penalties (Becker, 1968; Landes, 1983),¹⁴ emphasizing their deterrence properties, the more recent literature has focused on comparing penalty regimes in a second-best world. It is then assumed that, as is true in practice, penalties cannot be set so as to deter all or even most cartels.¹⁵ It is therefore important, in addition to the deterrence effect, to address the *price effects* of penalty regimes on cartels that are *not* deterred. An extensive recent comparison of the welfare properties of most of the penalty regimes described below is contained in Katsoulacos et al. (2015).

Clearly a penalty regime is better than another one if it is easier to implement, generates less uncertainty, and has a superior overall welfare impact. Unfortunately, regimes that are superior in terms of their welfare properties are not superior (and may in fact be inferior) in terms of other assessment criteria. This makes it difficult to translate results regarding the welfare properties of different regimes into proposals concerning which of these regimes should be adopted and implemented by CAs in practice.

Most CAs throughout the world have advocated for “simple”¹⁶ revenue-based monetary penalties for cartels.¹⁷ Many countries also ex-

9 The cost required in order to collect the necessary information and to undertake the estimation of the penalty by the CA and the firms.

10 The more the information required and the more difficult it is to obtain reliable data on this information the more lengthy will be the process of estimation and hence the greater the delay in reaching decisions.

11 That is, the cost for the CA of defending its decisions in Courts of Appeal and the cost that the firms have to incur when appealing against the CA's decisions.

12 Since penalties are generally calculated as a fraction of a “*penalty base*” (such as revenues or profits) and since penalty guidelines only specify that this fraction (the “*penalty rate*”) will fall within a range that will depend on a large number of mitigating and aggravating circumstances, there is always some uncertainty in predicting the CA's penalty estimate in any specific case. This uncertainty increases as it becomes more difficult to obtain accurate estimates of the “*penalty base*” and to calculate the appropriate “*penalty rate*”.

13 Thus, it may lead to under deterrence when lower penalties are mistakenly anticipated by potential offenders or over deterrence when innocent agreements are deterred by overestimating fines. Further, and very importantly in practice, the less *discretion* an agency has (limiting uncertainty) the less the degree of litigation on the amount of the fine by companies fined and the lower the risk of being accused of discrimination and public criticism of subjectivity and arbitrariness. See for details on this ICN Report (2008). ICN Report (2008) “Setting Fines for Cartels in ICN Jurisdictions” Report to the 7th Annual Conference, Kyoto, April 2008.

14 Becker (1968), “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 76, 169-217; Landes (1983), “Optimal Sanctions for Antitrust Violations,” *The University of Chicago Law Review*, 50, 652-678.

15 For example, due to bankruptcy considerations and in order not to violate the legal “*proportionality principle*” most countries have legal ceilings on antitrust fines set as a percent of annual turnover. These may well make penalties insufficient and antitrust policies either completely ineffective or at best partially effective in such a way that only low prices are deterred, while high prices are still sustainable. For details, see e.g. Buccirosi & Spagnolo (2007), *supra* note 8; Harrington (2010), *supra* note 8; Bageri et al. (2013), *supra* note 6; Katsoulacos & Ulph (2013), *supra* note 6; Houba et al. (2018), “Legal Principles in Antitrust Enforcement,” *The Scandinavian Journal of Economics*, 120(3), 859-893.

16 By “simple” we mean here not just that the penalty base (revenue) is easy to calculate but also and mainly that the penalty rate applied to the revenue is not related in a systematic way to the gravity of the specific offence.

17 See for example Bageri & Katsoulacos (2014), “A Simple Quantitative Methodology for the Setting of Optimal Fines by Antitrust and Regulatory Authorities,” *European Competition Journal*, Volume 10, 2014, 253-278. As noted in the ICN Report (2008) “the general view been that turnover/volume of affected

explicitly provide in their statutes for the imposition of penalties based on illegal gains (9 out of the 17 countries that participated in the ICN survey in 2008, including the US and China¹⁸). Penalties based on illegal gains can either take the place of revenue-based penalties (as in the U.S.) or they can be an additional penalty that is combined with the revenue-based penalty in order to reach the overall figure imposed on law violators (as in China). However, illegal gains-based penalties are rarely implemented – for example in U.S. in only one case has the imposition of a penalty based on illegal gains been implemented.¹⁹ Nevertheless, in some younger jurisdictions in which competition law is formulated in relation to the imposition of sanctions, CAs have opted for including illegal gains-based penalties as a potential additional element that can be taken into account when calculating monetary penalties.²⁰ However, this has not improved the implementation record of these illegal gains-based penalties due to the difficulties in their estimation and the uncertainty they create.²¹

Most of the literature on the optimal design of antitrust monetary penalties has focused on four main regimes: a damages-based regime, an illegal gains-based regime, a revenue-based regime and an overcharge-based regime. In a recent paper, (Katsoulacos et al., 2017), the authors also examine a fifth alternative regime, the *sophisticated revenue-based penalty regime*, in which the penalty *base* is the revenue of the cartel, but the penalty *rate* depends on (and increases with) the cartel overcharge rate. Finally, we should recognize that the literature on estimating private damage claims notes that this is based on a *simplified version of the damages-based penalty*. Therefore such a simplified version could also be used under public enforcement (by CAs). Finally, a *simplified version of an illegal gains-based regime*,²² has to be included for a full comparison. Below, we compare and contrast these seven potential penalty regimes.

commerce provides a good proxy for assessing the gravity of the behavior, both in terms of damage to consumers and illegal gain. Furthermore, such data is relatively easy to obtain” (p. 19).

18 See ICN Report (2008), p. 19. For an earlier detailed overview of the penalty structures implemented in OECD countries see the OECD Report (2002), “Fighting Hard-Core Cartels: Harm, Effective Sanctions and Leniency Programs”.

19 As we have been informed in a private communication with Greg Werden. This is generally true as also found in the survey of the ICN (Report 2008).

20 For example, on June 17, 2016, the Chinese CA enforcing law in the area of price-related anticompetitive conduct (“NDRC”) published “Draft Guidelines on the Determination of Illegal Gains and Fines in Relation to Undertakings’ Monopoly Conduct which are expected to be introduced formally this year. With these the NDRC attempts to make illegal-gains an important part of penalty setting in China - this has been commended by Wong-Ervin et al. (2016) – though in the past, the Chinese authorities have tried to calculate illegal gains in setting penalties in only about 10 percent of the cases. Also the Chilean Competition Authority (“FNE”) has adopted penalties based on illegal gains in 2014. See Wong-Ervin et al. (2016), “Monetary Penalties in China and Japan,” GMU Antonin Scalia Law School, DP 16 – 40.

21 “Difficult to be estimated” and “Easy to be challenged” is the standard way of explaining why penalties based on illegal gains are rarely used. Appendix 1 also provides more formal analysis to support these arguments. See, for the case of China, Deng & Katsoulacos (2017), “Anti-trust sanctioning in China: how can the NDRC guidelines be further improved?,” *Competition Policy International Antitrust Chronicle*, August 2017.

22 If a simplified version of a damages-based regime is admitted in the comparison, it is hard to justify not to include also a simplified version of the illegal gains-based regime.

II. BRIEF DESCRIPTION AND COMPARISONS OF THE MAIN MONETARY PENALTY REGIMES

Let us assume a market in which (p^c, Q^c) represent the cartel price and output while the (potentially imperfectly competitive) “but-for” price and output are (p^B, Q^B) and c is the marginal cost (“but-for” price under perfect competition). $R^c = p^c Q^c$ is the cartel revenue (turnover) while $R^B = p^B Q^B$ is the but-for revenue.

This is illustrated in Figure 1 below.

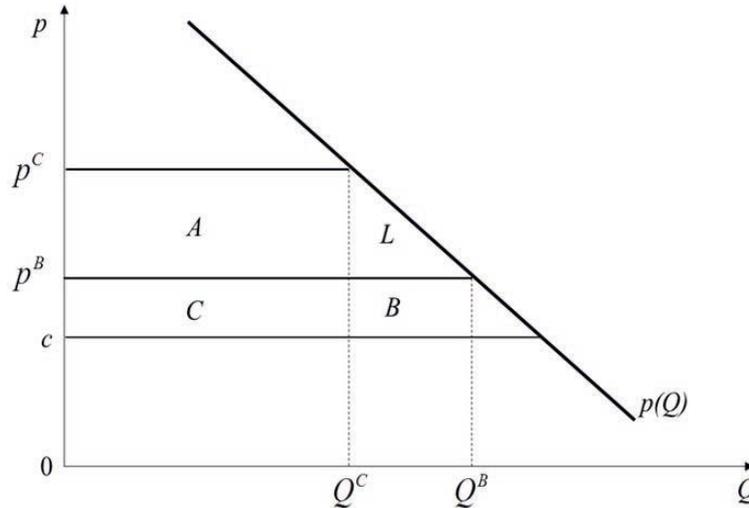


Figure 1

Let $\Delta\pi$ be the increase in profit if a cartel is formed and let the expected penalty be $\beta\phi B(p)$ where β is the probability that the cartel is detected, ϕ is the *penalty rate* and $B(p)$ is the *penalty base* – that is usually the cartel revenue, but can be a measure of the increase in profits or of the damages caused by the cartel. Generally, any anticompetitive action that increases profit for the firm taking it by $\Delta\pi$ will be undertaken depending on whether

$$\Delta\pi \begin{matrix} > \\ < \end{matrix} \beta\phi B(p)$$

The following penalty regimes (indicated by F_i), $F_i, i = D, G, R, O, SR$, have been discussed in the literature – and some have also been used in practice, to varying degrees.

1. Damages-based penalties (F_D). Generally, if an anticompetitive action causes damages (or harm) to others (D), a damages-based penalty is one for which:

$$\beta\phi_D B_D(p) = D$$

This ensures that when $\Delta\pi > D$ the action will be taken while if $\Delta\pi < D$ the action will not be taken. It is in this sense that total welfare is maximized with damages-based penalties. To achieve this outcome we can set the penalty rate $\phi_D = (1/\beta)$ in above equation and:

$$F_D = B_D(p) = D = (p^c - p^B)Q^c + L = \left(\frac{\theta}{1+\theta}\right) R^c + L \quad (1)$$

where $\theta = (p^c - p^B) / p^B$ is the proportional overcharge.

2. Illegal gains (or profit)-based penalties (F_G). With cartels, it is always the case that $\Delta\pi < D^{23}$ so a CA can use illegal gains-based penalties (illegal gains, $\Delta\pi$, being the cartel's profits over and above the counterfactual level of profits²⁴) to deter all cartels.²⁵ This requires that:

$$\beta\varphi_G B_G(\mathbf{p}) = \Delta\pi$$

which can be done by setting $\varphi_G = (1/\beta)$ and:

$$F_G = B_G(\mathbf{p}) = \Delta\pi = (p^C - p^B)Q^C - m\Delta Q = \left(\frac{\theta}{1+\theta}\right) R^C - m\Delta Q \quad (2)$$

where ΔQ is the reduction in output caused by the cartel and m is the absolute profit margin in the but-for situation, that is, the difference between but-for price and marginal cost.

Alternatively, the CA can use:

3. Revenue-based penalties (F_R). As already mentioned, these are the penalties most often adopted and implemented by CAs throughout the world. They are given by:

$$F_R = \beta\varphi_R B_R(\mathbf{p}) = \rho_R R^C \quad (3)$$

so penalty rate $\varphi_R = (\rho_R / \beta)$ and the penalty base $B_R(\mathbf{p}) = R^C$ (3')

In practice, the rate ρ_R falls within a range that depends on a large number of mitigating and aggravating circumstances.

4. Overcharge-based penalties (F_O). These are calculated as a multiple of the "but-for" revenue R^B . Specifically:

$$F_O = \beta\varphi_O B_O(\mathbf{p}) = \rho_O \theta R^B \quad (4)^{26}$$

so the penalty rate $\varphi_O = (\rho_O \theta / \beta)$ and the penalty base $B_O(\mathbf{p}) = R^B$ (4')

5. Sophisticated revenue-based penalties (F_{SR}). These use as their base the revenue obtained by the cartel (as in (3)), but the penalty rate depends on (and increases with) the cartel overcharge rate (as in (4)). Thus:

$$F_{SR} = \beta\varphi_{SR} B_{SR}(\mathbf{p}) = \rho_{SR} (\theta) R^C \quad (5)$$

so the penalty rate $\varphi_{SR} = (\rho_{SR} (\theta) / \beta)$ and the penalty base $B_{SR}(\mathbf{p}) = R^C$ (5').

Katsoulacos et al. (2017) examine in detail the simple case in which $\rho_{SR} = \theta$ and the penalty rate $\varphi_{SR} = \theta / \beta$.

23 Where $D = A+L$, refers to the *total consumer welfare harm* caused by the cartel price increase over the (counterfactual or) but-for competitive level, including, that is, the deadweight welfare loss triangle (L) associated with the reduction in the volume of output by the cartel.

24 In the special case where the counterfactual price is the marginal cost (competitive price), the illegal gains are the same as the cartel profits.

25 In terms of Figure 1, illegal gains are equal to area A-B.

26 In the special case in which the counterfactual / competitive price is the marginal cost (c) this is given by:

$$F_O = \rho_O \frac{Q(c)}{Q(c(1+\theta))} \left(\frac{\theta}{1+\theta}\right) R^C.$$

6. Simplified versions of damages-based and illegal gains-based penalties (\tilde{F}_D , \tilde{F}_G). Note that a “simplified” version of the damages-based penalty can be obtained by assuming that *damages, are estimated as in private damage claims*²⁷ - obtained by neglecting L in (1), so:

$$\tilde{F}_D = \tilde{B}_D(\mathbf{p}) = \tilde{D} = \left(\frac{\theta}{1+\theta} \right) R^C \quad (6)$$

with penalty rate $\varphi_D = (1/\beta)$.

Similarly, a simplified version of the illegal gains-based penalty can be obtained by neglecting ΔQ (the reduction in output), in which case the illegal-gains based penalty will be:

$$\tilde{F}_G = \tilde{B}_G(\mathbf{p}) = \tilde{\Delta\pi} = \left(\frac{\theta}{1+\theta} \right) R^C \quad (7)$$

again with penalty rate $\varphi_G = (1/\beta)$.

Table 1 shows the information required for calculating the above penalties. The information is categorized as Observable (O) or Unobservable (U) and in accordance with the difficulty in getting the information, as H: High, M: Medium and L: Low. This allows us to compare different penalty regimes in terms of the criteria of implementability and transparency.

Penalty (equation)	Damages-based (1)	Illegal gains-based (2)	Revenue based (3)	Over charge-based (4)	Sophisticated revenue-based (5)	Simplified damages- and illegal gains-based (6) and (7)
Information Required						
Cartel revenue $p^C Q^C = R^C$ (O; L)	X	X	X		X	X
Cartel volume of sales, Q^C (O; L)	X (for L)	X (for ΔQ)				
Counterfactual price and, hence, Overcharge $\theta = (p^C - p^B) / p^B$ (U; M)	X	X		X	X	X
Counterfactual volume of sales Q^B (U; H)	X (for L)	X (for ΔQ)		X		
Cost Information (c) (U; H)		X (for m)				
Information about Demand Structure (U; H)	X	X		X		

Table 1: Information required for the calculation of alternative penalties

Comparisons: Implementability and Transparency

As can be seen in Table 1, other than cartel revenue (which is the only information required by the simple revenue-based penalty regime) and the cartel's volume of sales, which are observable and obtainable at low cost, all other information required for implementing the other penalty regimes is unobservable and only obtainable at a medium to high cost. The following comments can be made regarding the properties of implementability and transparency:

²⁷ See Brander & Ross (2006), “Estimating Damages from Price-Fixing,” Canadian Class Action Review, 3(1): 335-369; Brander & Ross (2017), “Estimating Damages to Direct and Indirect Purchasers in Price-Fixing Actions,” Canadian Competition Law Review, Vol. 30, No. 1, 2017, pp 1-39.

Damages-based penalties: these include the deadweight welfare loss (L) and clearly are very difficult to estimate accurately. This is because the calculation of L requires knowledge of the but-for price and volume of sales and further information about the structure of demand. Thus, their estimation is likely to be subject to quite significant errors. Hence, such penalties have very significant implementability problems and a low degree of transparency, raising significantly the probability of them being challenged as false or discriminatory. For these reasons they very rarely form the basis for antitrust enforcement in practice.

Illegal gains-based penalties: these are also very difficult to estimate accurately through (2), as their estimation requires knowledge of the but-for price and volume of sales and hence about the structure of demand, as well as cost information (to estimate m). Thus, their estimation is likely to be subject to quite significant errors, which implies that such penalties also have significant implementability problems and can result in a low degree of transparency/significant amount of uncertainty.²⁸

(Simple) revenue-based penalties: as is clear from Table 1, these owe their popularity to the fact that they score high in terms of ease of implementation and also high on transparency (low uncertainty²⁹).

Overcharge-based penalties: their calculation is based on obtaining estimates of the price overcharge and, more importantly, the *counterfactual* volume of sales, in other words, information about the structure of demand.³⁰ This implies that this regime also scores low in terms of its implementability and transparency.

Sophisticated revenue-based penalties: these require for their calculation data on the cartel's revenue, as well as estimates of the price overcharge. Thus, these penalties score as “moderate” in terms of ease of implementation and transparency. *We should note that they have exactly the same ease of implementation and transparency as for obtaining standard estimates of damages in private damage claims.* We will elaborate on this in greater detail in the sections below.

Simplified damage-based and illegal gains-based penalties: Exactly the same remarks as for sophisticated revenue-based penalties apply here.

Thus we are led to the following:

Remark 1: The above discussion demonstrates that sophisticated revenue-based penalties are clearly superior when judged *in terms of ease of implementation and in terms of transparency (low uncertainty)* to the overcharge-based, the illegal gains-based, and the damages-based penalties. They are equivalent to the simplified damages-based and the simplified illegal gains-based penalties, although these do not perform as well, in terms of these criteria, as the simple revenue-based penalties.

Corollary to Remark 1: Competition authorities which, as we believe is true in reality, value ease of implementation and transparency over welfare impact will never adopt overcharge-based, or illegal gains-based or damages-based penalties because they score very low in terms of the two primary criteria.

The question is whether CAs might adopt one of the other penalty regimes (sophisticated revenue-based, simplified damages-based, or the simplified illegal gains-based), which are equally attractive in terms of ease of implementation and transparency, rather than the currently used simple revenue-based regime. Before answering this question we note that in their recent paper Katsoulacos et al. (2018a) show that, in terms of welfare impact, a linear sophisticated revenue-based penalty in the form of $F_{SR} = \rho_{SR}(\theta)R^C$ where $\rho_{SR}(\theta) = \theta$ is superior to the simplified damages-based or the simplified illegal gains-based regimes, as it leads to superior price effects.³¹ Thus, we have:

Remark 2: Sophisticated revenue-based penalties should be considered superior to the simplified damages-based or the simplified illegal

28 Nevertheless, because they are thought to have good deterrence properties, as already noted, they are sometimes included in the penalty regimes adopted, though they are very rarely implemented in practice. Concerning their welfare properties, it should be stressed that, in terms of price effects, they are inferior to overcharge-based and damages-based penalties (Katsoulacos et al. (2015), *supra* note 6; Katsoulacos et al. (2017), *supra* note 7.

29 On the other hand, as mentioned below, these penalties are very weak in terms of their welfare properties.

30 It is the product of these that forms the “penalty base” of this regime. Note that providing estimates of the overcharge (or the but-for price), as is done for private damage claims, is much easier than doing this *and* also providing an estimate of the but-for volume of sales, which, exactly in order to avoid the difficulties, is avoided in the standard approach used to calculate damages in private damage claims in which (6) is used. See also discussion on implementation below.

31 Katsoulacos et al. (2018a), “Cartel penalties and private damage actions: An integrated assessment,” mimeo, paper presented at CRESSE 2018 conference.

gains-based penalties because they are better in terms of welfare impact and equivalent in terms of ease of implementation and transparency.

This leaves only the comparison between the simple and the sophisticated revenue-based regimes.

Katsoulacos et al. (2017) show that a sophisticated revenue-based regime, given by (5) with $\rho_{SR}(\theta) = \theta$, is welfare superior in terms of *both* deterrence and price effects to a simple revenue-based regime in ensuring cartel prices below the monopoly level. Given that the sophisticated revenue-based regime is superior to the simple revenue-based regime in terms of both its price effects and its deterrence effects, it is likely that these beneficial effects outweigh any drawbacks in terms of ease of implementation and transparency and so we conclude that serious consideration should be given to switching the monetary penalty regime under public enforcement to a sophisticated revenue-based regime. We consider this recommendation in more detail in the next section.

III. WHY SHOULD COMPETITION AUTHORITIES SERIOUSLY CONSIDER SWITCHING TO THE SOPHISTICATED REVENUE-BASED PENALTIES?

Given the current state of knowledge, as reviewed in the previous sections, the only argument that can be used in order to justify the continued use of simple revenue-based monetary penalties in public enforcement against cartels, despite their poor welfare properties, is that by doing so we avoid the implementability and transparency problems associated with getting estimates on the overcharge. Here, we take a closer look at these problems.

The implementability and transparency concerns raised by the need to calculate the price overcharge, as under the sophisticated revenue-based regime, are very often vastly overstated. To explain why we consider this to be the case we note that the overcharge in cartel cases has been a magnitude that has been routinely estimated for many years in private damage claims to calculate damages, as given by expressions (6) or (7) above. These have been a very important feature of North American jurisdictions and have been introduced in EU competition policy since 2014, gradually becoming popular within EU countries too. It is now broadly recognized that there are many mature alternative methodologies for estimating the overcharge in damages claims that range from a low to a high degree of sophistication and so, as two prominent authors in this area wrote recently “Overall, we feel that a great deal of progress in damage estimation and related topics has been made in the past two decades. In addition, data availability has significantly improved and computing power has increased greatly. Therefore, good estimates of damages from price-fixing and related anticompetitive practices can often be obtained.”³²

An additional concern, often raised regarding the issue of having to calculate the overcharge in order to take it into account in setting monetary penalties, is that CAs would be overburdened if they became responsible for this. As the argument goes, in private damages claims those claiming damages undertake the estimation and the Courts just have to balance the evidence presented and choose between these and the counter estimates made by the defendants. However, a moment's thought indicates that this is certainly not a strong argument. This is because there is nothing to stop the CAs from requesting the parties (defendants and plaintiffs) to make available their own estimates of price overcharge, with detailed justification, *together* with all the other documents that they are asked to produce during the investigative procedure. Indeed, such a request, if mandatory, would likely have beneficial welfare effects since it would increase the costs of detection for cartel offenders – having to try to show low overcharge rates before this is required for dealing with private damage claims, and will incentivize plaintiffs not to make false claims of law violation. Of course, there will be cases where there are no claimants and the CA opens an investigation *ex-officio*. But in these cases too, it is certainly possible, as has been the standard practice in *ex-officio* investigations, for the CA to call on those that it recognizes as being harmed by the cartel and to request for them to provide evidence of the extent to which they were harmed – hence, of the overcharge rate. Clearly, these third parties will have ample incentives to provide this information since this will also be used in private damages claims.

³² Brander & Ross (2017), *supra* note 25. See also Brander & Ross (2006), *supra* note 25.

IV. CONCLUDING REMARKS

Taking into account the above arguments it seems very strange that, if paying damages is so widely accepted (as estimated in practice, using (6) or (7)) under private enforcement, there is opposition³³ to basing the calculation of monetary penalties on an expression like (5) - under public enforcement. One possibility for this opposition is that the case has not been adequately articulated in the past, taking into account all the considerations discussed above and, in particular, the welfare distortions caused by the currently used penalty regime as stressed in recent economic literature.

Another consideration that may be relevant concerns what the acceptable *burden of proof* is under private and public enforcement. To use the U.S. as an example, U.S. Courts have held that, while for claiming damages plaintiffs must show the existence of an injury with a “reasonable degree of certainty,” the proof of the amount of a plaintiff’s damages is subject to a *lower burden of proof* (*J Truett Payne Co v. Chrysler Motors*, 1981). The Supreme Court has held that damages may be shown using a “just and reasonable estimate, based on relevant data, including both “probable and inferential as well as direct and positive proof” (*Zenith Radio Corp. v. Hazeltine Research Inc.*, 1969). Thus, Courts have recognized the inherently lower ability to estimate damages and have accepted damage estimates based on reasoned analysis and partial information. Is there a reasonable reason why what is accepted by the Courts as burden of proof for private damages claims, should not or cannot be accepted by CAs? This is a legal rather than an economic question: should the burden of proof be higher for estimating penalties to punish and deter wrong-doing than for estimating them in order to compensate those that have been harmed by said wrong-doing? This point has nowhere been explicitly argued and justified. For as long as this remains the case, it does not seem possible to provide a convincing case for maintaining the current policy on monetary penalties in public antitrust enforcement.

³³ Beyond that associated with normal and to some extent understandable institutional inertia.



A COMPETITION LAW ANALYSIS OF COMMON SHAREHOLDINGS

BY NEIL CAMPBELL¹



¹ The author is a partner in the Toronto office of McMillan LLP. This article is a revised and condensed version of a paper presented at the 2018 CRESSE annual conference.

I. INTRODUCTION

Common shareholdings have begun to come under scrutiny in oligopolistic industries where multiple institutional investors own voting interests in multiple publicly-traded companies. Notable examples include studies of common shareholdings in the major airlines operating in the United States² and in some of the major banks in the United States and Europe.³ Issues related to the market power effects of common shareholdings in public companies were also discussed at some length in an annex to the European Commission's *Dow/Dupont* decision⁴ and were the subject of a recent roundtable session at the OECD Competition Committee ("OECD").⁵ However, relatively little attention has been paid to the application of existing competition law frameworks to such potential concerns.

As the OECD has observed, there has been a substantial growth in the types and activities of investment funds over the past half century.⁶ The holdings of individual institutional investors in common shareholder situations are typically less than 5 percent of the voting equity in any individual operating company, and they may well be as small as 1 percent or lower. Thus, they are much lower than the levels of cross shareholdings or other minority shareholder investments that have historically attracted competition law consideration (e.g. 10-25 percent, depending on the jurisdiction) and they usually are not accompanied by any right to representation on the board of directors.

Nevertheless, the cumulative interests of institutional investors with common shareholdings may be in the 15-25 percent (or above) range in multiple firms in some oligopolies, and they may be among the largest shareholders in some companies whose shares are widely held. For example, in *Dow/Dupont*, the Commission determined that "a small number of common shareholders, 17, collectively own around 21% of BASF, Bayer and Syngenta and around 29%-36% of Dow, Dupont and Monsanto."⁷

The theoretical literature, led by contributions from Elhauge⁸ and Posner *et al.*,⁹ has asserted that competitors who have significant shareholders in common will be incentivized to compete less aggressively with each other. The starting point is a variation on a "unilateral effects" theory of harm, but with a focus on common institutional shareholders, not on the firms themselves:

In the example of an oligopolistic market in which an institutional investor holds a minority of all (or most) firms, it is certainly theoretically possible that a unilateral price increase by one of the firms would be profitable from the investor's perspective. Losses from diversion of customers to competitors could be recouped because of the gains these competitors realize.¹⁰

The second step in the theory is that those common institutional investors will have the ability and incentives to induce the various firms they have invested in to raise their prices or otherwise compete less aggressively. The third step in the theory is that the firms that have such shareholders will choose to raise their prices or otherwise compete softly (in order not to damage those important shareholders' other interests – even though this may be contrary to the interests of remaining shareholders who do not hold shares in such competitors, and even though doing so may be a breach of fiduciary duties).

Coordinated effects theories of harm have also been extended to common shareholding situations. The potential basis for concern is that such linkages at the shareholder level could facilitate the reaching of understandings between the firms in which they hold shares to not compete

2 J. Azar, M.C. Schmalz & I. Tecu, "Anti-Competitive Effects of Common Ownership" (March 15, 2017), Ross School of Business Paper No. 1235 ("Azar et al., Common Ownership").

3 J. Azar, S. Raina & M. Schmalz, "Ultimate Ownership and Bank Competition" (July 23, 2016) ("Azar et al., Ultimate Ownership").

4 European Commission, *Dow/Dupont*, Case M.7932, Decision (March 27, 2017).

5 See OECD Secretariat, "Common Ownership by Institutional Investors and its Impact on Competition (Background Note)," DAF/COMP(2017)10 (November 29, 2017), available at [https://one.oecd.org/document/DAF/COMP\(2017\)10/en/pdf](https://one.oecd.org/document/DAF/COMP(2017)10/en/pdf) ("OECD Note").

6 OECD Note at para. 23.

7 *Dow/Dupont*, Annex 5, at para. 90.

8 E. Elhauge, "Horizontal Shareholding," 129 *Harvard Law Review* 1267 (2016); and E. Elhauge, "The Growing Problem of Horizontal Shareholding," *CPI Antitrust Chronicle* (June 2017).

9 E. Posner, F. Scott-Morton & E.G. Weyl, "A Proposal to Blunt the Anti-Competitive Power of Institutional Investors," forthcoming in *Antitrust Law Journal*.

10 OECD Note at para. 33. This is an extension of D. O'Brien & S. Salop, "Competitive Effects of Partial Ownership," 67 *Antitrust Law Journal* 559 (2000).

aggressively and/or could increase incentives to not deviate from coordinated outcomes.¹¹ If major competitors in an oligopoly each compete less aggressively, prices may end up above competitive levels and the firms may effectively exercise market power.

This paper explores how a standard merger review framework would apply to mergers in industries characterized by extensive common shareholdings, as well as the potential for review of acquisitions by institutional investors of common shareholdings in competing companies.

II. MERGERS WITHIN INDUSTRIES WITH COMMON SHAREHOLDINGS

Oligopoly theory does not provide clear benchmarks or methodologies to predict whether, and to what degree, oligopolistic competitors may engage in highly vigorous, versus moderate or soft competition. The common shareholdings literature suffers from the same lack of clarity. The intensity of competition in any particular oligopolistic market, with or without common shareholders, at any particular point in time requires a fact-specific, in-depth inquiry.

There is now widespread consensus in industrial organization economics that horizontal mergers between competitors give rise to economic welfare concerns when a transaction is likely to preserve or enhance the ability of firms to exercise market power. In general, merger reviews begin by defining relevant markets and then considering market concentration plus various other factors in order to assess competitive effects. When determining whether a merger would allow market power to be exercised, the critical comparisons are (i) the expected industry concentration and future behavior of the merging parties after, versus in the absence of, the merger; and (ii) whether other firms within or outside the relevant market are likely to respond in a manner that limits the ability of the merging parties to exercise market power.

A. Market Definition

The literature on common shareholdings focuses on concentrated oligopolistic industries. Relatively little attention is paid to the difference between the broad concept of an “industry” or “sector” and the much narrower approaches used to define relevant markets as a starting point for assessing whether or not market power can be exercised. Broadly defined oligopolistic industries will often involve numerous specific products and geographic areas. When assessing a merger between firms with common shareholders, there is no basis for abandoning the assessment of market power at the level of the relevant markets in which specific suppliers compete to sell to specific products or services to specific customers.

A relevant market may or may not include every firm that has common shareholders in a concentrated oligopoly, and may or may not include various other firms that will be important for a correct assessment of market power in a specific market. For example, the six major agro-chemical companies that were the focus of the common shareholdings analysis in *Dow/Dupont* were each involved in numerous relevant product and geographic markets, but there were some markets where not all were active, and there were various markets in which additional competitors were present.¹²

B. Competitive Effects

Market power — and the effects of a merger on the ability to exercise market power — is difficult to measure in practice. Factors likely to affect the possibility of market power being exercised include:

- *Market Shares and Concentration* — The common shareholdings literature makes considerable use of the Modified HHI (“MHHI”) measure that has been derived from the theoretical work of O’Brien and Salop.¹³ However, as with the traditional HHI measure of concentration, there is no exact MHHI level, or merger-induced change in the MHHI, that reliably indicates whether or not market power can be exercised. While the MHHI measure may provide directional indications regarding mergers that warrant in-depth review, it would not be appropriate as a “likely challenge” or “presumption of anti-competitive effects” threshold. It is also important to recognize that calculation of MHHIs can be an extremely resource-intensive exercise, since it requires detailed information regarding all the shareholders (including affiliation relationships between them) of all the competitors in each relevant market.

11 OECD Note at paras. 39-41.

12 *Dow/Dupont*, at parts V.4 – V.6.

13 O’Brien & Salop, *Partial Ownership*; *Dow/Dupont*, Annex 5, at part 6.2.

- *Competitive Vigor of the Party Being Acquired* – This factor (which is sometimes framed in terms of whether one of the merging parties is a “maverick” that inhibits coordinated market power effects) can be relevant in an analysis that takes common shareholdings into account, depending on the pre-merger ownership of the firm being acquired. If there are significant common shareholdings between the acquirer and other competitors in a relevant market, the common shareholdings theory would predict that the acquirer would not be a vigorous pre-merger competitor. A merger is unlikely to enable the incremental exercise of market power if one of the merging parties is not providing significant competitive discipline in the market.
- *Effectiveness of Remaining Competitors* – The ability of merging parties to exercise market power depends in large part on whether current competitors are likely to discipline a post-merger attempt to increase prices (or to reduce non-price dimensions of competition) or whether they are likely to accommodate and follow such an action by the merging parties. The extent of common shareholdings between the merging parties and each competitor in a relevant market is theoretically a relevant consideration in assessing whether that competitor is likely to attempt to undercut or to accommodate a post-merger attempt to exercise market power.
- *Buyer Power* – Merging parties’ customers may have sufficient countervailing power to resist an attempt to exercise market power (although it is important to consider whether such power is applicable broadly across the customer group or only to one or more specific purchasers). If the relevant buyers are publicly-traded companies, it is theoretically possible that common shareholdings between the merging parties and such buyers could affect the likelihood of an attempt being made to exercise market power (or, potentially, the likelihood that such buyers would respond by way of threatening or exercising their countervailing buyer power).¹⁴
- *Entry and Other Supply Responses* – Even if the competitors to the merging parties or their customers are not likely to discipline attempts to exercise market power, various types of supply-side responses may do so. Such responses may include firms which are not close competitors to the merging parties adopting repositioning strategies, firms in adjacent product or geographic markets expanding into the relevant market, greenfield entry by a new firm, and innovation that will affect future competition in the market place: All of these supply-side responses may in theory be affected by the degree of common shareholdings between firms that have the ability to undertake such supply responses and the merging parties (as well as other existing competitors in the market in question). This may be a resource-intensive inquiry.
- *Mechanisms by which Accommodating Responses Occur* – In evidence-based merger review systems, a general assertion of the theory that common shareholdings may increase the likelihood of unilateral or coordinated effects is unlikely to be sufficient to justify a remedial order in regard to a specific merger transaction. Evidence-based decision-making should require that the competition authority have some basis for concluding that:
 - (i) the merged firm would implement the price increases that are expected to be unilaterally profitable from the perspective of the institutional common shareholder(s); and
 - (ii) existing competitors, customers with buyer power and firms that are in a position to provide a supply-side response would choose to accommodate a post-merger exercise of market power rather than respond competitively.

The common shareholdings literature hypothesizes three main mechanisms by which the incentives of institutional investors with common shareholdings to prefer soft competition may be implemented and acted upon by the management of the companies in which they have invested:¹⁵

- *Shareholder Votes* – While institutional investors can influence company management through *shareholder votes*, these typically involve broad governance matters (e.g. elections of directors, appointments of auditors, approvals of major strategic transactions) and rarely involve day-to-day or even major issues of competitive strategy.
- *Direct Communications* – Meetings or other direct communications with senior management and/or directors of one or more competing firms are a less transparent mechanism than *shareholder votes* and could allow for more detailed communications regarding significant aspects of competitive strategy. However, this mechanism involves potentially significant risks to both the institutional investors and the firms if the discussion leads to understandings between such investors and the managers of two or more firms that

¹⁴ There is no *a priori* reason why common shareholdings would only be taken into account in respect of competitors but not suppliers or customers.

¹⁵ See OECD Note at paras. 51-80.

could constitute a “hub-and-spoke” conspiracy with potentially significant penalties under the competition/antitrust laws in various jurisdictions. In addition, senior management and directors would expose themselves to allegations of breaching their duties to act in the best interests of all their company’s shareholders, rather than just the sub-set of institutional shareholders which have the common shareholdings.

- *Compensation Incentives*— Compensation systems which encourage management personnel to focus on industry profitability instead of the company’s own profitability could lead to accommodating rather than competitive responses (if the major firms in an oligopoly all followed a similar approach). However, such systems could raise potential fiduciary duty issues for company directors and officers. In addition, management’s theoretical incentives could well be in conflict with similar incentives to not injure publicly traded customers and/or suppliers whose shareholders include common institutional investors.

The common shareholdings literature generally focuses on the interactions between senior management and shareholders. However, if competition agencies raise common shareholding theories of harm in actual cases, the merger review process will likely also need to consider whether there is any evidence of the transmission mechanism by which the senior management, who are assumed to be responding to the interests of common institutional shareholders, would implement instructions or incentives to compete softly to the relevant lower-level managers responsible for pricing and other dimensions of competition in specific relevant product and geographic markets.

C. Causation

The appropriate approach for evaluating the competitive effects of a merger is to compare the likely levels of prices and/or non-price dimensions of competition if the merger occurs against the levels that would likely prevail in its absence (i.e. a “but-for analysis”). Under such an approach, the focus is on whether the merger is expected to result in some preservation or enhancement of market power, relative to the non-merger scenario.¹⁶ However, if an industry has significant pre-merger common shareholdings, it may be extremely difficult to establish that there has been material incremental change in:

- the ability or incentives of institutional investors to influence management to compete more softly (or that management will choose to do so on its own), and/or
- the likelihood that other competitors and potential supply responders would adopt accommodating responses when they have been competing vigorously.

The common shareholdings literature predicts that such an ownership structure would already have led to such behavior by the common shareholders, the merging parties, the other competitors, and the potential supply responders, and that it would be expected to continue absent the merger. It would therefore be necessary to focus on whether the nature or extent of the unilateral anti-competitive behavior and/or the accommodating behavior would increase materially as a result of the merger, and thereby facilitate a greater degree of market power being exercised than was already occurring pre-merger.

In order to conduct a reliable assessment of the extent to which common shareholdings are likely to result in less competitive behavior by individual firms, considerable evidence regarding the existing behaviors of these firms is required. More specifically, this would appear to require competition authorities to undertake potentially time-consuming and burdensome processes to gather evidence regarding:

- the extent to which the merging parties were exerting meaningful competitive discipline on each other pre-merger, notwithstanding their partial common shareholder populations;
- the likelihood that the remaining publicly-traded competitors with partial common shareholders are going to become more accommodating than they were pre-merger;
- the likelihood that one or more firms that were going to reposition/expand/enter/innovate in the absence of the merger will choose not to do so as a result of the merger; and/or

¹⁶ The future non-merger scenario is often approximated by the pre-merger market conditions, although such an approach is only valid if it is reasonable to expect that the future is likely to resemble the past.

- the extent to which there are competitors, buyers with countervailing power, or supply responders, without common shareholders, that could provide competitive discipline on attempts to exercise an incremental level of market power post-merger.

The European Commission did not fully address these issues in *Dow/Dupont*. In Annex 5 of the decision, it set out a detailed summary of the theoretical and empirical literature about the potential market power effects of common shareholdings (albeit with little of the critical commentary that has emerged in response). However, due to procedural issues arising in the Statement of Objections process, it concluded that it could not base its decision on MHHs and the common shareholdings analysis.¹⁷ Instead, it commented that “common shareholding in the agro-chemical industry is to be taken as an element of context in the appreciation of any significant impediment to effective competition that is raised in the Decision.”¹⁸

In doing so, the Commission avoided undermining its own primary theories of harm. Had it placed greater reliance on the degree of industry concentration attributed to common shareholdings, it would have introduced significant internal contradictions into its competitive effects analysis:

- On one hand, the main basis for the Commission’s conclusion that there would be anti-competitive effects in various agro-chemical markets was that Dupont and Dow were important current sources of competitive discipline on each other, and also that they were important competitors in innovation and new product development.¹⁹
- Yet at the same time, the Commission asserted that the degree of common shareholdings between the merging parties and the other major global agro-chemical competitors was substantial.²⁰
- No explanation was provided as to how the aggressive current competition findings and the expected innovation/new product competition findings could be reconciled with the significant common shareholdings between Dow and Dupont (and other competitors). If the common shareholdings analysis was sound, the pre-merger competitive environment would already be characterized by soft competition, with market power being exercised by the merging parties and the other major competitors identified in that analysis.

In summary, where there are significant common shareholdings between the merging parties and their major competitors in an oligopolistic industry, the common shareholdings literature predicts that the merging parties would not be providing aggressive competitive discipline on each other’s prices or other non-price dimensions of competition. Thus, any finding of anti-competitive harm would have to be based on a determination that the elimination or reduction of some small amount of pre-merger competition nonetheless meets the applicable materiality standard in the legal test for challenging a merger transaction. The counter-intuitive implication for merger review processes is that, in general, there will be less of a reason to be concerned about a merger when there are already significant common shareholdings in an industry than mergers between competitors where there are little or no pre-existing common shareholdings.

III. INVESTMENTS BY INSTITUTIONAL SHAREHOLDERS

A. Reviewability of Acquisitions of Small Shareholdings

As the OECD has noted, many jurisdictions do not subject minority shareholdings to merger control, and those that do typically have a minimum voting share threshold (e.g. 15-25 percent) or a material influence threshold.²¹ Thus, acquisitions of small shareholdings such as the 1-5 percent levels of concern in the common shareholdings literature are not usually subject to review.

The addition of a notification requirement for these levels of share ownership (e.g. 1-5 percent, or possibly lower, voting interests of individual shareholders) would represent a major expansion of the scope of merger control in any such jurisdiction. Even if the regular financial (usually turnover) thresholds were maintained, lowering the minimum shareholding to the level where such small voting shares interests are acquired could generate reviews for an enormous volume of capital markets transactions. There is no easy way to limit the notification require-

¹⁷ *Dow/Dupont*, Annex 5, at paras. 74-79.

¹⁸ *Ibid.*, at para. 81.

¹⁹ *Ibid.*, at parts V.6 and V.8.

²⁰ *Ibid.*, Annex 5.

²¹ OECD Note, at paras. 7-15.

ment to common shareholdings in competing firms, let alone to concentrated oligopolistic industries, because these concepts are not definable in objective, administrable filing thresholds.²² Such reviews would have significant resource consequences for the reviewing agency, not just the institutional investors and the companies they are investing in.

Designing a notification regime for acquisitions of small voting equity positions would also give rise to a number of other challenges. For example:

- (i) Would transactions that have already been completed all be grandfathered?
- (ii) What level of incremental share purchases would be caught (e.g. suppose an investor proposes to increase its stake in a firm from 3 to 4 percent)?
- (iii) How would review processes be applied to day-to-day buying and selling of shares in public markets (i.e. to transactions which are not subject to take-over bid timing rules and which would be incompatible with even short no-close review periods)?

B. Substantive Review

The relevant markets for analysis of a new share purchase by a particular institutional investor would be those in which the company whose shares are being acquired competes with one or more of the other companies in which that same shareholder already has investments. As noted above, since most merger control regimes define relevant markets more narrowly than the “oligopolistic industry” concept discussed in the common shareholdings literature, there may be numerous relevant markets and they may be characterized by varying degrees of common shareholdings between the merging parties and other competitors, as well as varying levels of competition from privately-held or other firms that do not have any common shareholders.

With regard to competitive effects, any single acquisition of a new common shareholding by a single institutional investor may have relatively little impact on overall industry concentration, even under MHHI measures that are designed to reflect the aggregate concentration resulting from common shareholdings.

It will be similarly important to assess whether or not other competitive effects factors (e.g. impact on the behavior of the target, effectiveness of the remaining competition, countervailing buyer power and the likelihood of supply-side responses) are impacted in any material way by a particular acquisition of a small shareholding. For example:

- *Impact on the Behavior of the Target Firm* – If there already are other common shareholdings between the target firm and other competitors in the industry, will the emergence of an additional common shareholder have any material incremental effect on the target firm’s pricing or other competitive behavior? Alternatively, if the investment constitutes the first time that the target firm becomes subject to a common shareholding, will the position of this one institutional investor be significant enough to impact the target firm’s behavior?
- *Effectiveness of Remaining Competition* – If significant existing common shareholdings are already leading to an exercise of market power in the market, will the additional common shareholding have any material effect? Alternatively, if market power is not already being exercised as a result of common shareholdings, what is it about the new common shareholding that would be expected to result in market power being exercised in the future?
- *Buyer Power* – Does the new common shareholding materially reduce the ability or incentives of customers to exercise countervailing buyer power?
- *Supply-side Responses* – Does the new common shareholding materially reduce the likelihood that potential supply-side responders would discipline attempts to exercise market power?

In summary, in many cases the issues of causation and materiality of changes resulting from any single acquisition of a common shareholding are likely to make it very difficult for competition authorities to conclude that a purchase of a small voting interest by a single institutional

²² Clear and objective filing thresholds are vitally important to provide certainty for parties contemplating transactions and their advisors, as well as for competition agencies: See International Competition Network, *Recommended Practices for Merger Notification and Review Procedures*.

shareholder would meet the statutory test for an anti-competitive merger. In addition, it is far from clear that this type of complex intervention would be a good use of scarce enforcement resources.

C. Regulatory Approaches

Any approach which focuses on individual transactions on a go-forward basis will be limited in its effectiveness because of the vast pre-existing common shareholdings by multiple institutional investors in some concentrated oligopolies. There would also appear to be a degree of arbitrariness from focusing only on small future transactions.

Posner *et al.* have proposed that individual investors not be allowed to hold more than 1 percent of the market share in an oligopolistic industry (measured indirectly by summing their proportionate interest in the market share of each competitor they are invested in), unless:

- (i) they restrict their shareholdings to a single effective competitor in the industry (defined so as not to include small fringe players), or
- (ii) they operate only as a “passive investor” (which is defined to require (1) no communications at all with company management, (2) mirror voting of their shares in proportion to the voting of other shareholders so that they have no influence in any corporate governance decision (which is tantamount to not exercising their voting rights), and (3) committing “to own and trade stocks only in accordance with clear non-discretionary public rules, such as matching an index as closely as possible”).²³

Posner *et al.* contend that such a regime could be implemented by the US antitrust enforcement agencies through the issuance of policy guidance.²⁴ However, they acknowledge that antitrust authorities would have to prepare and regularly update lists that define oligopoly industries. Leaving aside potentially serious substantive, jurisdictional and transitional dislocation issues, their proposal is premised upon the ability of the US antitrust agencies to challenge completed transactions that fall below the HSR filing thresholds and to challenge past transactions without any limitation period restrictions. Such an approach would not be viable in most other jurisdictions.

As the OECD notes, several critiques have been expressed about this type of “hard limit on common ownership,” including that:

- current evidence is not sufficient to justify this type of *per se* rule;
- such a rule might effectively require large investment management firms to split up, resulting in costs and inefficiencies in capital markets;
- there would be significant monitoring and compliance burdens for investment firms and enforcement agencies;
- the proposed limits are overbroad and would interfere with many investments which do not result in actual anti-competitive effects; and
- limiting the exercise of voting rights in order to qualify for the passive investor exception could have negative corporate governance and capital markets implications.²⁵

At this point in time, the scope of the alleged competition policy problem from common shareholdings and the effect on overall economic welfare is not clear. The extent of the time and resources burdens — for merging parties, other market participants and competition authorities — are also not easy to quantify, but have the potential to be very significant. More comprehensive evidence of actual harm across a range of markets, as well as detailed cost/benefit analysis and consideration of broader implications for the operation of capital markets, would be important to determine whether extending merger control regimes or introducing regulatory constraints on small institutional shareholder investments would be an appropriate intervention in capital markets.²⁶

²³ Posner *et al.*, *Power of Institutional Investors*, at pp. 33-34.

²⁴ *Ibid* at pp. 34-35. Posner *et al.* note that legislative amendments or the issuance of rules/regulations are alternatives to enforcement agency guidance. In this author’s view, major economic policy shifts of this nature should be designed and implemented using legislation, regulations or rule-making after shareholder consultation, rather than through administrative discretion.

²⁵ See OECD Note at para. 106 and sources cited therein.

²⁶ For further discussion regarding the importance of corporate governance, capital market, financial regulatory and other broader policy considerations, see

IV. CONCLUDING OBSERVATIONS

The burgeoning common shareholdings literature has cast the spotlight on the possibility of sub-optimal levels of competition between public companies in concentrated oligopolies. However, it is important to differentiate between soft competition and exercises of market power which result from the phenomenon of common shareholdings, and those which result from conscious parallelism in oligopolistic industry structures.

Existing merger control frameworks will likely be of only limited utility for identifying and remedying mergers where competition concerns arise from the presence of common shareholdings (and where a traditional analysis focused on the merging parties would not have done so) or problematic incremental acquisitions of common shareholding positions by institutional investors. In both contexts, issues of causation and materiality are likely to limit severely the situations in which competition authorities could appropriately take action based on objective evidence and analysis. Moreover, the resource costs for agencies, merging parties and third parties are likely to be significant, casting doubt on whether the benefits of such reviews would exceed the costs. In an environment of scarce enforcement resources, even greater caution is warranted as many enforcement agencies are likely to have other areas where they can make larger contributions to improving economic welfare.

Instead, heightened awareness of the possible application of competitor agreement laws — and the development of more rigorous compliance programs — within both institutional investors and public companies may be one of the most important practical steps that can be taken to reduce anti-competitive influences on company management. In addition, it would be useful to establish a corporate governance principle that management compensation systems should be focused on company, rather than industry, performance. Major institutional investors should consider endorsing such a principle and encouraging all the companies in which they invest to implement it.

OECD Note at part 5.2 and para. 136.



FIVE NOT SO EASY PIECES TO MAKE ANTITRUST WORK FOR INNOVATION

BY RICHARD GILBERT¹



¹ Emeritus Professor of Economics, University of California, Berkeley.

I. INTRODUCTION

The movie “Five Easy Pieces” chronicles the struggles of a young man who abandons the microcosm of classical musicians in which he was raised to find his way in a hardscrabble world. The protagonist has many misadventures and the movie ends on a note of profound uncertainty about his future. In a not entirely dissimilar sense, antitrust enforcement matured in a microcosm of neoclassical price theory and is struggling to find its way in the hardscrabble world of high technology. Whether antitrust policy will become an effective tool to promote innovation is also as yet uncertain.

The U.S. Congress established the Antitrust Modernization Commission (“AMC”) to consider how antitrust law and enforcement can best serve consumer welfare in the global, high-tech economy that exists today and to assess whether existing antitrust laws are up to that task.² After three years of staff effort and many weeks of hearings, the Commission issued a Report which recognized that antitrust enforcers should carefully consider market dynamics in assessing competitive effects, but concluded that “[t]here is no need to revise the antitrust laws to apply different rules to industries in which innovation, intellectual property, and technological change are central features.”³

At one level these are not controversial findings. There is no disagreement that innovation is a key determinant of consumer welfare and that antitrust policy should be cognizant of the effects of mergers and acquisitions and firm conduct on innovation incentives. There is no need to revise the U.S. antitrust laws because they are sufficiently vague to apply flexibly to most any industry circumstance. They do not define competition, harm to competition, or monopolization and allow broad scope for courts to take idiosyncratic features of industries into account including the effects of conduct and market structure on industry dynamics and consumers.

Nonetheless, the AMC was unjustifiably sanguine in its conclusions because high technology industries have characteristics that are difficult to address following generally accepted antitrust principles. Antitrust enforcement developed over more than a century to embrace neoclassical price theory with its emphasis on short-run allocative efficiency, which applauds measures that move prices closer to marginal production costs. While this evolution has had positive effects for consumer-friendly enforcement in “old economy” industries, it limits the ability of courts to enforce antitrust laws to promote innovation in high technology “new economy” industries. These features include: the relative importance of competition to create new products and services compared to price competition; a typically large gap between marginal cost and average cost, which makes pricing at or near marginal cost a questionable objective; the potential for innovation to disrupt markets, which severs links that might otherwise connect existing market structures to future competition; conflicts between the need to protect information to encourage its creation and the social value of disseminating information widely; the role of network effects and platforms; and other complications such as standards, complementors, and interoperability that can be difficult to address following historical precedents.

My focus in this note is on merger policy for new economy industries, which I define as industries for which the development of new products or technologies is a critical strategic instrument.⁴ I identify five obstacles (“not so easy pieces”) that hinder sound antitrust enforcement for mergers in new economy industries.

II. LIMITED ECONOMIC THEORY AND EMPIRICAL EVIDENCE

There is not a large body of theoretical and empirical evidence that addresses the impact of mergers or firm conduct on innovation. This might surprise students of law and economics because there is a very large economic literature on competition and innovation.⁵ The problem for merger policy is that relatively few of these studies focus on mergers and a merger is not the same as a reduction in competition. Absent economics of scale or scope, a merger does not change the resources available for research and development, although it can change incentives to apply those resources to develop a new product or technology. In contrast, a reduction in competition removes R&D resources from an industry, which can have different effects for R&D incentives compared to a merger.

2 *Antitrust Modernization Commission: Report and Recommendations*, April 2007 at 1.

3 AMC Report at 9.

4 These are typically industries with high research and development (“R&D”) intensities as measured by the ratio of R&D expenditures to sales.

5 Philippe Aghion & Jean Tirole label the competition-innovation relationship the second most studied topic in the field of industrial organization, second only to the price-competition relationship. See Philippe Aghion & Jean Tirole, *The Management of Innovation*, 109 *Quarterly J. Econ.* 1185, 1195 (1994).

Several recent theoretical studies develop a foundation to evaluate how a merger changes the unilateral incentives⁶ of the merging parties to invest in R&D.⁷ These studies generally find that mergers lower unilateral investment incentives for R&D, absent efficiencies, spillovers or other industry characteristics that limit the ability of innovators to profit from their discoveries. Mergers create “downward innovation pressure,” much as they create “upward pricing pressure.”⁸ However, these studies have numerous limitations and have not been validated with empirical evidence.

Most of the recent theoretical studies that address merger incentives for innovation adopt a highly simplified description of industry dynamics. Typically, they assume only two time periods: an initial period in which firms invest in R&D and a subsequent period in which they compete in prices. A two-period model abstracts from important industry dynamics that can affect innovation incentives. For example, innovation incentives can differ greatly for firms that are close to or far from a technological frontier, a proposition that is supported by economic theory and empirical evidence and which can have implications for merger effects.⁹ These asymmetries are typically absent in two-period models, which typically assume that firms have symmetric innovation capacities in the R&D stage and ignore asymmetries that may develop when firms have different payoffs from their research and development efforts.

In addition to the complications from industry dynamics, simple models abstract from efficiencies that may lower the cost or increase the effectiveness of R&D expenditures and from spillovers or other industry characteristics that affect the ability of innovators to profit from their discoveries. Efficiencies and appropriation benefits can reverse predictions that mergers exert downward innovation pressure.¹⁰ More research is needed to identify the circumstances under which these effects are likely to be quantitatively significant.

III. ANTITRUST’S OBSESSION WITH MARKET DEFINITION

It did not take long for courts charged with enforcing the antitrust laws to require identification of relevant markets to evaluate allegations of monopolization or attempted monopolization.¹¹ In 1911 the Supreme Court interpreted the Sherman Act’s prohibition of monopolization to have “both a geographical and distributive significance.”¹² The 1914 Clayton Act prohibited acquisitions whose effects may be to constrain commerce in any section or community or tend to create a monopoly in any line of commerce.¹³ The phrase “line of commerce” invites market definition.

Following William Baxter, mergers can raise distinct concerns in three different categories of market activity. They can have concerns for “today’s products” that are presently being delivered in the marketplace, the activity of R&D, and “tomorrow’s products,” the goods and services

6 Unilateral incentives refer to incentives that do not assume responses by non-merging parties. See U.S. Department of Justice and Federal Trade Commission, *Horizontal Merger Guidelines*, August 19, 2010 at § 6.

7 See, e.g. Igor Letina, *The Road Not Taken: Competition and the R&D Portfolio*, 47 *RAND J. Econ.* 433 (2016); Michael A. Salinger, *Net Innovation Pressure in Merger Analysis*, available at <https://ssrn.com/abstract=3051249> (2016); Massimo Motta & Emanuele Tarantino, *The Effect of Horizontal Mergers, When Firms Compete in Prices and Investments*, UPF Working Paper No.1579, August 30 (2017); Giulio Federico, Gregor Langus & Tommaso Valletti, *Horizontal Mergers and Product Innovation: An Economic Framework*, 59 *Int. J. Industrial Org.* 1 (2018); and Ángel L. López & Xavier Vives, *Overlapping Ownership, R&D Spillovers, and Antitrust Policy*, *J. Political Economy* (forthcoming).

8 For a description of “upward pricing pressure,” see Joseph Farrell & Carl Shapiro, *Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition*, 10 *B.E. J. Theoretical Econ.: Policies and Perspectives* 1 (2010) and the *Horizontal Merger Guidelines* at § 6.1.

9 See, e.g. Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith & Peter Howitt, *Competition and Innovation: An Inverted-U Relationship*, 120 *Q. J. Econ.* 701 (2005); Philippe Aghion, Richard Blundell, Rachel Griffith, Peter Howitt & Susanne Prantl, *The Effects of Entry on Incumbent Innovation and Productivity*, 91 *Rev. Econ. & Stat.* 20 (2009); and Nicholas Bloom, Mirko Draca & John Van Reenen, *Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity*, 83 *Rev. Econ. Stud.* 87 (2016).

10 See, e.g. Vincenzo Denicolò & Michele Polo (2018), *Duplicative Research, Mergers and Innovation*, 166 *Econ. Letters* 56 (2018) for an example of R&D efficiencies from mergers. For a discussion of merger appropriation benefits, see, e.g. Vincenzo Denicolò & Michele Polo, *The Innovation Theory of Harm: An Appraisal*, (March 22, 2018), available at <https://ssrn.com/abstract=3146731>; Bruno Jullien & Yassine Lefouili, *Horizontal Mergers and Innovation* (August 1, 2018), available at <https://ssrn.com/abstract=3135177>; Marc Bourreau & Bruno Jullien, *Mergers, Investment and Demand Expansion*, 167 *Econ. Letters* 136 (2018); and Richard Gilbert, *Competition, Mergers and R&D Diversity*, (October 8, 2018), available at <https://ssrn.com/abstract=3190478>.

11 For a discussion of the evolution of market definition in antitrust cases, see, e.g. Gregory J. Werden, *The History of Antitrust Market Delineation*, 76 *Marq. L. R.* 123 (1992).

12 *Standard Oil Co. v. United States*, 221 US 1, 61 (1911).

13 Enacted October 15, 1914, codified at 15 U.S.C. §§ 12–27, 29 U.S.C. §§ 52–53.

that may eventually be delivered as a consequence of successful R&D.¹⁴ The traditional approach to market definition addresses only the delineation of markets for “today’s products.” With respect to the activity of R&D, the *Antitrust Guidelines for the Licensing of Intellectual Property* define a “research and development market,” which “consists of the assets comprising research and development related to the identification of a commercializable product, or directed to particular new or improved goods or processes, and the close substitutes for that research and development.”¹⁵ However, with the narrow exception of contract research and development, R&D is not a traded product or service.¹⁶

The effects of mergers on “tomorrow’s products” also can be difficult to address within the conventional contours of market definition. Suppose it is likely that a merger of firms Alpha and Beta will reduce R&D investment and likely delay the production of a new technology to identify genetic defects. There may be no market for the technology if it does not presently exist or its future boundaries may be difficult to assess. If there is a discernible market for the technology, the effect of the merger may be to delay output in that market, which is arguably different from lessening competition in an existing or future market.

At least one court has opined that a firm cannot be held liable for antitrust harm in a market that does not presently exist. In *SCM Corp. v. Xerox Corp.*, a district court heard a complaint that Xerox engaged in various practices that excluded SCM from markets for plain and coated paper copiers, including acquisitions of patents. While noting that patent acquisitions are not exempt from Section 7 of the Clayton Act, the judge concluded that “Liability for retrospective money damages cannot be predicated under § 7 upon a patent acquisition made prior to the existence of a relevant product market.”¹⁷ Even if *SCM v. Xerox* is not the law of the land, market definition is not a particularly helpful tool to evaluate harm to innovation. It is a challenge to apply a hypothetical monopolist test to a market that does not yet exist. It is difficult to translate a “small but significant and non-transitory increase in price” to conduct that delays a new technology.

The fact that market definition is difficult, and perhaps impossible, for innovation does not mean that antitrust enforcement should ignore innovation. Market definition can obscure, rather than inform, the analysis of competitive effects in high technology industries for which existing market structures are highly imperfect predictors of future competition. Louis Kaplow, an accomplished scholar of law and economics, wrote that market definition is “impossible and counterproductive” even for existing goods and services because “there does not exist a valid means of inferring market power from market shares” and “it is impossible to determine which market definition is superior without already formulating one’s best estimate of market power, rendering the exercise pointless.”¹⁸ The 2010 revision of the U.S. Department of Justice and Federal Trade Commission *Horizontal Merger Guidelines* subtly suppresses the role of market definition in agency guidance by transposing the order of “market definition” and “competitive effects,” placing the latter before the former and adding that “The Agencies’ analysis need not start with market definition.”¹⁹

Perhaps the historically central role of market definition in antitrust analysis is not fatal for the evaluation of harms to innovation. But the antitrust Agencies will have to develop new techniques to evaluate these harms and may face opposition from courts that expect the sophisticated approaches to product and geographic market definition that have been developed for “today’s” products and services.

IV. STANDARD OF PROOF TO ESTABLISH ANTITRUST HARM

Courts increasingly require antitrust plaintiffs to empirically demonstrate alleged antitrust harms. In merger cases this may take the form of an econometric analysis relating prices to market competition as the Federal Trade Commission did in its 1997 challenge to the merger of two office supply superstore chains, Office Depot and Staples.

A key issue in the analysis of the proposed Office Depot/Staples merger was whether the superstores (Office Depot, Staples, and OfficeMax) were more effective competitors for some types of office supplies than other retailers, such as Target and Wal-Mart. By identifying

14 William F. Baxter, *The Definition and Measurement of Market Power in Industries Characterized by Rapidly Developing and Changing Technologies*, 53 *Antitrust L.J.* 717, 717-718 (1984).

15 U.S. Department of Justice and Federal Trade Commission, *Antitrust Guidelines for the Licensing of Intellectual Property*, January 12, 2017 at § 3.2.3.

16 At least one U.S. court rejected a market for R&D because the complaint did not identify one or more product markets consisting of reasonably interchangeable goods. See *Golden Gate Pharmacy Services, Inc. v. Pfizer, Inc.*, No. C-09-3854 (N.D. Cal. Dec. 2, 2009), *aff’d* 433 Fed. Appx. 598 (May 19, 2011).

17 463 F. Supp. 983, *aff’d* 645 F.2d 1195, 1210.

18 Louis Kaplow, *Market Definition: Impossible and Counterproductive*, 79 *Antitrust L.J.* 361, 361 (2013).

19 *Horizontal Merger Guidelines* at 7.

geographic pricing zones with different numbers of superstores and other retailers, the FTC could distinguish the effects of competition from superstores from competition from other retailers and concluded that the merger would significantly increase prices in many geographic regions.²⁰

It is unlikely that innovation harms can be predicted with a similar level of precision. Unlike retail competition, R&D competition is not geographically specific. It would be difficult, and possibly misleading, to identify regions of the country where merging firms compete in R&D and differentiate them from other regions where they do not compete in R&D. Research and development efforts may be local but its effects for innovation are potentially global. Even if local or temporal differentiation in R&D efforts might exist, it would be challenging to identify its effects. Innovation is inherently uncertain and difficult to measure. Common indicators, such as R&D expenditures and patent counts, are only indirectly related to innovation.

Without a way to calculate the effect of industry competition on innovation, antitrust enforcers are likely to resort to surrogate indicators, such as corporate documents that discuss R&D intentions. But corporate documents are imperfect predictors of future conduct and are susceptible to strategic manipulation if corporate executives appreciate how their records might influence enforcement outcomes. Empirical studies, buttressed by sound economic theory, which measure the effects of historical merger activity in an industry on innovation outcomes are likely to be an important information source if antitrust authorities pursue enforcement actions that turn on innovation effects. As noted, the empirical record is weak and much more should be done to accumulate relevant knowledge to support enforcement actions.

Effective antitrust enforcement for innovation will require courts to balance the costs of over-enforcement and under-enforcement. Plaintiffs bear the burden to prove their allegations and it is socially costly to hold parties liable for conduct that is unlikely to create harm. However, it would be unwise to abandon merger enforcement for innovation merely because effects cannot be predicted with a high degree of certainty. Decisions not to challenge mergers that may harm innovation are also subject to error and can have high resulting social costs.

V. TREATMENT OF EFFICIENCIES AND APPROPRIABILITY

While mergers can reduce the unilateral incentives of the merging parties to invest in R&D, they can have other impacts that promote innovation. Mergers can have efficiency benefits that enable the merged firm to conduct R&D more effectively. They can eliminate redundant R&D expenditures, allowing the merged firm to reduce costs and perhaps re-direct R&D effort to other productive activities. The merged firms also may have complementary assets and technological expertise that can increase the productivity of R&D. Furthermore, a merged firm can benefit from intra-firm spillovers that allow each merger partner to profit from R&D efforts by the other partner. In addition, a merger can allow the merged firm to appropriate greater value from its discoveries than the merging parties can realize individually by increasing price and output in markets for new products and services and by limiting informational spillovers that allow rivals to profit from its discoveries.

The AMC Report recommended that the FTC and the Antitrust Division of the DOJ should give substantial weight to evidence demonstrating that a merger will enhance consumer welfare by enabling the companies to increase innovation.²¹ The Horizontal Merger Guidelines acknowledge that mergers can have efficiency benefits for research and development. However, the Guidelines also state that “Research and development cost savings may be substantial and yet not be cognizable efficiencies because they are difficult to verify or result from anticompetitive reductions in innovative activities.”²² Of course, antitrust authorities should recognize efficiencies only if they are verified and cost-savings do not offset anticompetitive effects if they are integral with harm to competition. However, courts historically have been reluctant to accept efficiency claims as a defense in merger cases and there is little justification to impose a higher threshold of proof for R&D efficiencies.²³

The Horizontal Merger Guidelines add that “The Agencies also consider the ability of the merged firm to appropriate a greater fraction of the benefits resulting from its innovations,”²⁴ but the Guidelines do not detail how mergers may increase appropriation or where the burden of

20 See Orley Ashenfelter, David Ashmore, Jonathan B. Baker, Suzanne Gleason & Daniel S. Hosken, Empirical Methods in Merger Analysis: Econometric Analysis of Pricing in *FTC v. Staples*, 13 *Int. J. of Econ. of Bus.* 265 (2006).

21 AMC at 10.

22 HMG at 31.

23 William J. Kolasky & Andrew R. Dick, The Merger Guidelines and the Integration of Efficiencies Into Antitrust Review of Horizontal Mergers, 71 *Antitrust L. J.* 207, 209 (2003). (Courts have been slow to accept efficiencies as an integral part of the competitive effects analysis of mergers).

24 HMG at 31.

proof lies to establish appropriation effects. A merger can increase appropriation by increasing the profits from an innovation²⁵ or by internalizing spillovers that would otherwise dilute innovation incentives. The profit from an innovation can increase if the merger increases the future price or output of the innovation. Higher prices enhance welfare only if the positive effects from greater innovation (or higher quality) more than compensate for the consumer harm from higher prices. Greater output is a consumer benefit, but the mechanism that promotes higher output needs to be explicated because, all else equal, higher prices tend to lower output.

Spillovers have complicated effects for merger incentives. Spillovers allow rivals to imitate discoveries without bearing all of the costs of R&D. These external spillovers reduce the ability of innovators to profit from discoveries and therefore lower innovation incentives. Mergers can promote innovation by reducing the number of rivals that can imitate a discovery. Spillovers that are internal to the merged firm allow each merger party to benefit from discoveries made by the other party. These internal spillovers can allow the merged firm to achieve greater innovation benefits than each party could achieve on its own or allow the merged firm to amortize the cost of R&D over a larger output.

A key question is who bears the burden to evaluate appropriation benefits from a merger, including the effects from external and internal spillovers. The merging parties correctly have the burden to demonstrate cost-saving efficiencies because they have more information than antitrust enforcers about the ability of a merger to lower their costs, including R&D expenditures. It is unclear whether the merging parties have greater information about the ability of a merger to achieve appropriation benefits. The parties may have given little thought to quantifying appropriation benefits in their normal course of business and arguably antitrust enforcers could have greater competence to explain how a merger would or would not facilitate appropriation. The ability to mount a successful appropriation defense depends on the assignment of the burden to justify appropriation benefits, and courts have not resolved this issue.

VI. EFFECTIVENESS OF REMEDIES TO RESTORE INNOVATION INCENTIVES

The most common outcome of a merger challenge brought by an antitrust agency is a consent decree in which the merging parties agree to abide by specified structural or behavioral conditions. These may include some combination of asset divestitures, supply commitments, and technology licensing obligations. There is an economics literature that addresses whether these consent agreements have prevented price increases for mergers and acquisitions.²⁶ The U.S. FTC and other enforcement agencies have conducted their own retrospective analyses of merger remedies,²⁷ but they have not examined whether consent decrees have restored innovation incentives for cases that alleged innovation concerns. It is difficult to restore innovation incentives that are lost from a merger. A successful divestiture requires sufficient capital, labor, technical experience, and intellectual property rights for the recipient of the divested assets to replace lost R&D, along with the will to apply these assets to innovate in technical areas that the merging parties would have pursued if they did not merge.

Spinning off a division of the merged company to a third party may not be sufficient to achieve these goals. Antitrust authorities cannot order employees of the merging companies to move to the designated recipient of the divested assets. Skilled personnel may prefer a different employer, location, or occupation. One study of acquisitions in the pharmaceutical industry found that only 22 percent of the employees that were listed as inventors on the acquired company's patents moved to the acquiring company.²⁸ Furthermore, the merged company might retain intellectual property rights or exclusive relationships that impose barriers to successful innovation by the recipient of the divested assets.

If further study demonstrates that partial remedies often fail to remedy alleged innovation harms from mergers, the antitrust authorities can refuse to accept partial remedies and instead unconditionally challenge mergers that threaten innovation. However, the threat of an unconditional challenge can cause further harm. Suppose firms Alpha and Beta plan to merge and have overlapping R&D activities in one of their many divisions. If they anticipate that a reviewing agency will move to block the merger because the agency cannot design an effective remedy, they may "fix it first" and spin off the troublesome division to a third party. Of course, there is no reason to expect that the fix orchestrated by the merging parties would be any more effective than a remedy negotiated with the reviewing agency.

25 The relevant measure is the incremental profit from invention relative to the profit if a firm does not invent. A merger can increase the incremental profit from invention under some circumstances even if it does not increase the gross profit from invention.

26 See, e.g. John Kwoka, Daniel Greenfield & Chengyan Gu, *Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy*, MIT Press (2015).

27 See, e.g. U.S. Federal Trade Commission, *The FTC's Merger Remedies 2006-2012: A Report of the Bureaus of Competition and Economics*, January 2017.

28 Colleen Cunningham, Florian Ederer & Song Ma, *Killer Acquisitions* (August 28, 2018), available at <https://ssrn.com/abstract=3241707>. This percentage would not indicate a loss of future innovation competition if patentees that went elsewhere more than compensated for the loss of innovation by the acquiring company.

The solution to the problem of ineffective remedies to address innovation harms is better information about the types of remedies that work and those that fail. Studies should also explore the effectiveness of different types of remedies, such as compulsory licensing of intellectual property as an alternative to asset divestitures.

VII. THE ROAD AHEAD

The Report of the AMC concluded that “No substantial changes to merger enforcement policy are necessary to account for industries in which innovation, intellectual property, and technological change are central features.”²⁹ At a very general level, this conclusion is correct. But this note explains that courts and antitrust enforcers have a lot of work ahead to make merger enforcement an effective policy tool to promote innovation.



²⁹ AMC Report at 9.

CPI Subscriptions

CPI reaches more than 20,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles, and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.

