



Volume 9 | Number 2 | Autumn 2013

Competition Policy International

ANTITRUST AND INTELLECTUAL PROPERTY

Robin Jacob, Timothy Wu, Douglas Ginsburg & Joshua Wright, and Richard Epstein & David Kappos on striking the right balance and general principles

Herbert Hovencamp on consumer welfare and IP law

Eliana Garcés-Tolón and Kai-Uwe Kühn on antitrust, FRAND, and SSOs

Pierre Régibeau and Fiona Scott-Morton on reverse payment settlements

CURRENT CASES

Julie Bon, Pietro Crocioni & Francesca Sala on Anglo/Lafarge

THE CLASSICS

Levin, Klevorick, Nelson, and Winter's *Appropriating the Returns from Industrial Research and Development* with an Introduction by **David S. Evans**



Edited by David S. Evans

FOUNDER AND PUBLISHER

David S. Evans: *University of Chicago Law School, Chicago, IL, USA*

EDITOR-IN-CHIEF

David S. Evans: *University of Chicago Law School, Chicago, IL, USA*

CO-EDITORS

Elisa V. Mariscal: *Centro de Investigación y Docencia Económicas (CIDE), and Instituto Tecnológico Autónomo de México (ITAM) Mexico*

Mark Armstrong: *Oxford University, Oxford, UK*

Antonio Bavasso: *Allen & Overy, London, UK*

Keith N. Hylton: *Boston University School of Law, Boston, MA, USA*

R. Ian McEwin: *Chulalongkorn University, Bangkok, Thailand*

Susan Ning: *King & Wood, Beijing, China*

Michael A. Salinger: *Boston University, School of Management, Boston, MA, USA*

CHAIRMAN

Richard Schmalensee: *MIT Sloan School of Management, Cambridge, MA, USA*

EDITORIAL BOARD

Joseph Angland: *White & Case, New York, NY, USA*

Dennis W. Carlton: *University of Chicago, Chicago, IL, USA*

Einer R. Elhauge: *Harvard Law School, Cambridge, MA, USA*

Nicholas J. Forwood: *General Court of Appeals of the European Union, Luxembourg*

Herbert Hovenkamp: *University of Iowa College of Law Iowa City, IA, USA*

Valentine Korah: *University College London London, UK*

Mario Monti: *Bocconi University Milan, Italy*

Eduardo Pérez-Motta: *Federal Competition Commission and International Competition Network Mexico City, Mexico*

Patrick Rey: *IDEI and University of Toulouse, Toulouse, France*

Christopher Bellamy: *Linklaters London, UK*

Frank H. Easterbrook: *US Court of Appeals for the 7th Circuit, Chicago, IL, USA*

John Fingleton: *Fingleton Associates, London, UK*

Douglas H. Ginsburg: *US Court of Appeals for the DC Circuit, Washington D.C. USA*

William J. Kolasky: *Wilmerhale, Washington D.C. USA*

Bruno Lasserre: *Autorité de la Concurrence Paris, France*

Timothy J. Muris: *George Mason University, Fairfax, VA, USA*

Richard Posner: *US Court of Appeals for the 7th Circuit Chicago, IL, USA*

Vivien Rose: *Competition Appeal Tribunal, London, UK*

Bo Vesterdorf: *Plesner and Herbert Smith Freehills, Copenhagen, Denmark; Brussels, Belgium*

ON THE WEB

www.competitionpolicyinternational.com

Table of Contents

Introduction

Antonio Bavasso and David Evans ix

Symposium on Antitrust and IP

Competition Authorities Support Grasshoppers: Competition Law as a Threat to Innovation

The Rt. Hon. Professor Sir Robin Jacob 15

Intellectual Property Experimentalism by Way of Competition Law

Tim Wu 30

Whither Symmetry? Antitrust Analysis of Intellectual Property Rights at the FTC and DOJ

Ginsburg and Wright 41

Consumer Welfare in Competition and Intellectual Property Law

Herbert Hovenkamp 53

Legal Remedies for Patent Infringement: From General Principles to FRAND Obligations for Standard Essential Patents

Richard A. Epstein and David J. Kappos 69

Licensing of Standard Essential Patents: Antitrust Intervention is Not Big Enough a Fix

Eliana Garcés Tolón 90

Table of Contents *(continued)*

Justifying Antitrust Intervention in ICT Sector Patent Disputes: How to Address the Hold-Up Problem

Kai-Uwe Kühn 100

“Pay-for-Delay”: What Do We Disagree On? Jevons Seminar, Washington, DC, November 5, 2013

Pierre Régibeau 116

Pay-for-Delay

Fiona Scott Morton 128

Current Cases and Issues

There is Always a First Time: Coordinated Effects via Vertical Structural Changes in Anglo/Lafarge

Julie Bon, Pietro Crocioni and Francesca Sala 139

The Classic

Appropriating the Returns from Industrial Research and Development

Richard C. Levin, Alvin K. Klevorick, Richard R. Nelson, and Sidney G. Winter 157



Introduction

Introduction

BY ANTONIO BAVASSO AND DAVID EVANS

The CPI Autumn 2013 issue features contributions from some of the participants to two roundtables on the intersection between intellectual property and antitrust that we organized in London (in June 2013) and Washington, DC (in November 2013) with the Jevons Institute for Competition Law and Economics of University College London. The theme remains controversial and, during the two roundtables, participants exchanged strong and diverging views under the Chatham House rule. This divergence of views, a veritable intellectual divide, is well reflected in the articles published in this issue, although perhaps not the intensity of the heat.

Let's start with the common ground: both antitrust and intellectual property (IP) law “are concerned with promoting economic welfare.” Both systems of law aim at providing incentives for innovation. But they do so from fundamentally different vantage points. IP law has evolved from a policy judgment in that the way to promote innovation is, in some cases, to provide inventors with property rights that could allow them to capture a return from their efforts and prevent others from encroaching on their creations—for some time and with some limitations. Antitrust law is concerned with establishing rules for the game of competition that prevent businesses from rigging the outcome in ways that exclude competition and harm consumers.¹ Neither approach is necessarily inconsistent with the other but, then again, neither approach ensures that either is consistent with the other.

Herbert Hovenkamp's article looks at the common grounds and at the divergences that flow from the welfare standards that typically inform competition policy enforcement. He concludes that implanting a consumer interest consideration onto the IP system would require a major overhaul of the system, and it is not clear that this is needed or is the solution to the perceived problems.

Antitrust has the ability to introduce consumer welfare considerations in relation to unilateral behavior of the IP right holder or indeed any agreement relating to the IP right. But is that interference or oversight required or warranted? On the one hand, Tim Wu appears to welcome a certain degree of “legal experimentalism.” He regards the Supreme Court's *Actavis* judgment as an endorsement of “the use of competition laws to examine the potential excesses of the current patent regime.” On the other hand, Robin Jacob rejects competition law's interference in his forceful article. He starts with the well-known fable of the grasshopper and the ants and concludes that competition authorities help grasshoppers (i.e. those who only look for profits for today and do not contribute to the future) “positively helping them sing in the summer” (i.e. gather the profits now) rather than the ants (i.e. those who use part of their income of today to build for the future). He looks specifically at pharmaceuticals and telecoms and concludes by asking competition authorities to “leave the ants alone” and let the patent system do the job that it was designed to do.

A key aspect of the intellectual divide relates to the question of whether there is something special about the characteristics of intellectual property which means that general rules on property or antitrust should not apply. Epstein and Kappos, in their piece, look at legal instruments under common law and patent law, at the

legal dynamic on contract formation and breach, negotiation post breach, and damages and injunctions in patent cases, including recent litigation under FRAND standards in relation to standard setting organizations (SSO). They recognize that there is a risk of hold up between contracting parties and hold out with strangers. They also recognize the danger of strategic behavior from deliberate breaches but conclude that those dangers rest with both parties, not just one. They therefore think that it would be “unwise to tailor liability rules on the assumption that either a landowner or a patent holder is filled with guile, while an actual or potential trespasser or infringer acts with purely virtuous motivations.” They would leave these issues for contract law to sort out generally.

Patents, Epstein and Kappos note, are probabilistic nature of patent—that is, until it is challenged, and kicked around by a finder of fact, it is not clear whether a patent really is worth the paper it is printed on. Securing a patent is not like buying a farm. This characteristic has been explored in the work of Shapiro and others over the last decade and has formed a strong basis for the antitrust intervention. The relevance of this characteristic features prominently in the articles by Fiona Scott-Morton and Kai-Uwe Kühn that focus on, respectively, the pharmaceutical and the information and communication technology (ICT) sectors.

In relation to the pharmaceutical sector, the pieces by Fiona Scott Morton and Pierre Regibeau should be read in tandem. They both address the question of pay-for-delay in light of the *Actavis* judgment while also taking into account the different regulatory regime in the US than in the European Europe. Scott Morton is far more skeptical of reverse payment settlements than Regibeau.

With regard to the ICT sector, Kai-Uwe Kühn claims that—from an economic perspective—there is a strong and convincing body of theory indicating that, as a result of the probabilistic nature of patents, hold-up is likely to occur. His view is that the objection that hold-up needs to be anchored in a proof of hold-up in individual cases ignores the importance of the counterfactual and would make antitrust enforcement not viable.

He recognizes that these theories sit better within a legal system, such as the European one, which outlaws abuse of a dominant position and accommodates exploitative abuses rather than a monopolization standard such as Section 2 of the Sherman Act.

In essence, as explained in Kühn’s piece, the policy that emerges from the ICT cases on the use of injunctions by SEP patent holders that are FRAND encumbered is as follows: the alleged harmful effect (the hold up) is strongly grounded in theory and therefore a behavior that fits that pattern is *per se* unlawful under Article 102. In some respects the *per se* nature of this approach could be seen as turning an *ex post* enforcement tool into a form of *ex ante* regulation.

Many have been worried about the impact that this might have on the attempt to introduce an “effects based” standard to unilateral practices abuses, particularly Article 102 enforcement.

Moreover this approach—which is limited to the specific issues of SEP agreed in the context of SSO—raises an important question of principle as to whether there is something unique about those cases that requires a different form of presumption. The article by Ginsburg and Wright raises that question mostly from a US law perspective, but similar reconsideration would apply in other legal regimes. They conclude that antitrust enforcement in relation to intellectual property and other property rights should be based on a principle of

symmetry, including in relation to presumption with respect to use of injunctions with regard to FRAND encumbered SEPs.

Finally, the piece from Eliana Garcés Tolón argues that antitrust intervention is not a big enough fix. Some would argue that that is because it is not the right fix. Eliana Garcés Tolón does not take issue with the question of whether the current theoretical approach side-steps the effects analysis of proving the harm caused by each individual case of alleged abuse. Rather, she notes that antitrust authorities deliberately avoid getting involved in FRAND determination, and they limit their intervention to arguing that the value attributable to hold-up cannot be taken into account in determining the FRAND value as that would be abusive. She notes that antitrust policy has probably neared the limit of its reach and SSO can help clarifying the boundaries of FRAND terms with the help of the courts.

In fact, if the role of public antitrust enforcement is to indicate that any hold-up value should be discounted in FRAND determination, an adjudicator or a court would be much better suited to resolve that point by interpreting a FRAND standard as shown most recently by Judge Holderman's opinion in the *Innovatio* case. Antitrust (consumer or total welfare) considerations do not add much to the notion of fairness and/or reasonableness that is intrinsic in the FRAND standard. But the European Commission brings the question of FRAND back into the frame (albeit indirectly): it takes issue with seeking injunctions against a "willing" licensee and it is difficult to see how a licensee can be characterized as willing unless one determines whether his offer was truly FRAND.

The different positions are quite well delineated on the battleground but the controversy is likely to remain. Shall we go to Court? The CJEU is on the case.

The editors of CPI have had their own little saga with intellectual property rights in the weeks leading up to this issue. As dedicated readers know, in every issue we publish a classic article—often an important one that is seldom read anymore but should be. For this issue, we thought an interesting one would be Fritz Machlup and Edith Penrose's "The Patent Controversy in the Nineteenth Century."² They wrote their article at a time, around 1950, when there was considerable debate over reforming the patent system. They wanted to remind the world that much of the debate was a redo of similar heated discussions that occurred during the 19th century. And that economic theory and empirical evidence were no more advanced in the mid 20th century than it had been a century before. Machlup, in particular, took the view that there was not enough economic support to adopt a patent system, but given that we had one, not enough to dismantle it either. It is a very nice article. Unfortunately, for the first time in 9 years, we have been unable to reach an agreement to reprint a classic. Cambridge University Press said no. But if we have piqued your curiosity, you can buy it³ or go to the University of Texas Law Review⁴, which has apparently snubbed its nose at copyright or had better luck with Cambridge.

Our failed negotiation with Cambridge University Press led us, however, to a much younger but still exceptional work. In 1987, Richard Levin, Alvin Klevorick, Richard Nelson, and Sidney Winter published a pioneering empirical study of how companies appropriate the returns for research and development. Surprisingly, they found that companies used multiple methods to protect their returns and that patents, while taken, were not viewed as important safeguards of returns in some industries. Their research added a great deal to our empirical knowledge on the role and use of patents and is still frequently cited today. Unfortunately, in our view at least, this article did not change Machlup and Penrose's 1950 conclusion concerning the lack of solid

empirical information to guide the policy debate over the proper role of patents.

Finally, to add some counterweight to the patent debate we have included, for our regular deep dive into a case, an article by Julie Bon, Pietro Crocioni & Francesca Sala on the merger of the UK operations of Anglo and Lafarge. The UK Competition Commission blocked the merger based on concerns over coordinated effects arising from vertical aspects of the combination of these entities, which were involved in the production of both cement and RMX. The case is unusual as the authors point out because it is rare to block a merger because of coordinated effects and rarer still because of vertical considerations.

Thanks to all of the authors for their excellent contributions. And to the Brookings Institution for allowing us to reprint the Levin et al. article. ▲

1. Evans, David and Keith Kylton, "The Lawful Acquisition and Exercise of Monopoly Power and Its Implications for the Objectives of Antitrust," 4(2) *Competition Policy International* (Autumn 2008).
2. Machlup, Fritz, and Edith Penrose, "The Patent Controversy in the Nineteenth Century," *The Journal of Economic History* 10 (1950): 1-29.
3. <http://www.jstor.org/discover/10.2307/2113999?uid=3739256&uid=2&uid=4&sid=21103143317301>
4. [http://www.utexas.edu/law/journals/tlr/sources/Volume 92/Issue 2/Hemel/Hemel.fn002.machlup.thepatentcenturyin-nineteenthcentury.pdf](http://www.utexas.edu/law/journals/tlr/sources/Volume%2092/Issue%202/Hemel/Hemel.fn002.machlup.thepatentcenturyin-nineteenthcentury.pdf)



*Symposium on
Antitrust and IP*

Competition Authorities Support Grasshoppers: Competition Law as a Threat to Innovation

BY THE RT. HON. PROFESSOR SIR ROBIN JACOB¹

Aesop's fable is of the ant and the grasshopper. In summer the ant works, gathering and storing food against the coming winter. The grasshopper laughs and sings, living only for the day. Innovators are the ants: using part of the income of today to build for the future. The grasshoppers are those who look only for profits for today and do not contribute to the future. Competition authorities are favouring grasshoppers, positively helping them sing in the summer (i.e. gather in profits now) and saying, "do not worry, sing away. When winter comes we will make the ants feed you." The danger to innovation is all too obvious for those who can see: it pays to be a grasshopper rather than an ant—better to be a copyist than an innovator.

I shall demonstrate this by reference to the competition authorities' conduct in relation to two industries heavily dependent on innovation: pharma and telecoms. Let us recall some basics.

I. BASICS: IP LAW ITSELF HAS INBUILT REGARD TO THE NEED FOR COMPETITION

IP rights are exceptions to the general Western model of free competition. They are justified by the advantages they provide to society outweighing the advantages provided by free competition. Some types of IP are completely uncontroversial. Thus, no-one contests that the law should prevent one trader from falsely representing his goods or services to be that of another. He may compete with another, but not by lying to the public. Where trade mark law goes further—for instance by preventing a man from honestly comparing his goods or services with another—controversy begins.² But the debate takes place within the context of the nature of the IP right itself; competition is taken into account in the legal assessment of the extent of the right. Competition law just does not come into.

As regards patent law its very basic rules are built round the idea of free competition. You cannot patent that which is old or obvious. Why not? Because otherwise, you would interfere with free competition. Nor can you patent more than you have invented. Why not? Because you would monopolise that which should be free for others to explore. A patent is limited by the rather crude 20-year³ term. Why term-limited? Because a longer monopoly would interfere with free competition.

COMPETITION AUTHORITIES ARE FAVOURING GRASSHOPPERS, POSITIVELY HELPING THEM SING IN THE SUMMER (I.E. GATHER IN PROFITS NOW) AND SAYING, "DO NOT WORRY, SING AWAY. WHEN WINTER COMES WE WILL MAKE THE ANTS FEED YOU." THE DANGER TO INNOVATION IS ALL TOO OBVIOUS FOR THOSE WHO CAN SEE: IT PAYS TO BE A GRASSHOPPER RATHER THAN AN ANT—BETTER TO BE A COPYIST THAN AN INNOVATOR.

Thus, patent law rights and concepts are riddled with non-interference with competition. What you can patent is the practical application of new and ingenious ideas. You get a monopoly but it is a monopoly in something

which would not have existed (or existed as soon) but for your inventive contribution. The law interferes not with ordinary competition, but with competition in something which would not have existed but for the inventor.

The availability of a future monopoly is a major driver for innovation. It is no good bemoaning this reality.

YOU GET A MONOPOLY BUT IT IS A MONOPOLY IN SOMETHING WHICH WOULD NOT HAVE EXISTED (OR EXISTED AS SOON) BUT FOR YOUR INVENTIVE CONTRIBUTION. THE LAW INTERFERES NOT WITH ORDINARY COMPETITION, BUT WITH COMPETITION IN SOMETHING WHICH WOULD NOT HAVE EXISTED BUT FOR THE INVENTOR.

It is no good advocating abolition and replacement of the patent system with some alternative, such as State-sponsored research rewards. It has been tried (for instance, in the Soviet Union) and has not worked. There is, of course, much government or charity supported research, particularly in the field of medicines. But the reality is that the patent system remains the bedrock of future research both for new medicines and new medical devices.⁴ Abolitionist or dilutionist economists have been around a long time—since the mid-19th century, at least. All the current talk of thickets, “hold-up” and the like has happened before, and fortunately largely⁵ seen off with the result that innovation proceeds faster and faster.

II. THE APPROACH OF COMPETITION AUTHORITIES

Given that patent law is so aware of competition, why do competition authorities wish so much to restrict patent rights further? Or, what amounts to the same thing, the enforcement or attempted enforcement of patents? I think there are several, interrelated, reasons.

First is that they are largely staffed by somewhat theoretical economists. They not only have a very imperfect understanding of how the patent system actually works but also similar lack of understanding of how business—particularly innovative business—works and is financed. Patent monopolies do not fit their models of competition. These models look to the present, not the future. Of course, current patent monopolies interfere with current competition. They make current prices higher than they would be in a state of perfect competition—some economists’ ideal market. That some patents cover important products which do not have any significant alternative makes them worse in these economists’ eyes because such patents command markets in “inelastic goods”—something for which there is no substitute.

This view easily gains ill-informed populist support. Politicians always jump on a bandwagon when they see one. Who is not in favour of lower prices? Anyone who stands in the way of these—such a patentee—is a bad guy. But deep down, what the competition authorities are doing is pushing for instant gratification in the shape of lower prices⁶ to consumers now at the expense of the benefits of delayed gratification in the shape of innovation for the future. I shall demonstrate this by reference to two of the most important subject matters of patent protection: pharma and telecoms. In both of these industries the competition authorities have weighed in heavily in favour of the copyists and against the inventors. In so doing they are in conflict with other agencies (see, for instance, the recent highly pro-IP report from OHIM⁷) who support innovation and in conflict with innovating companies. They play the part of the Mayor of Hamlyn town. I quote from Browning’s poem:⁸

“If I can rid your town of rats
Will you give me a thousand guilders?”
“One? fifty thousand!” – was the exclamation
Of the astonished Mayor and Corporation.

Later, when the rats had all been drowned, the Mayor said:

“So, friend, we’re not the folks to shrink
From the duty of giving you something to drink,
And a matter of money to put in your poke;
But as for the guilders, what we spoke
Of them, as you very well know, was in joke.
Beside, our losses have made us thrifty.
A thousand guilders! Come, take fifty!”

III. THE PHARMA SECTOR INQUIRY

There was a preliminary report in 2008 and a final report in 2009.⁹ What led to the inquiry in the first place, I do not know. But the manner in which it was conducted was a disgraceful use of the Commission’s powers. For it began with simultaneous dawn raids on major pharma companies throughout Europe. Dawn raids without a judicially issued search warrant and without any justifiable suspicion that evidence would be destroyed are strongly reminiscent of a lawless regime. A Commission official at a conference I was at recently described dawn raids as “normal!” That itself is shocking to anyone who cares about civil liberties.

The preliminary report was appalling. It revealed a vast ignorance on the Commission’s part as to how the patent system actually works. In introducing it the then-Commissioner, Neelie Kroes, on 28th November 2008 said:

Several of the most damaging practices which delayed or blocked market entry of competitors include:

1. *Patent clustering*, where a company forms a dense network of patents around a medicine. The worst example we found of this method was 1300 separate patent filings, across the EU, for a single medicine.
2. A *large number of litigation cases* over patents, which originator companies invoked against generic companies. On average, these cases took three years to resolve, and originator companies lost a clear majority of cases.
3. *Patent settlements* which constrain market entry of generic companies, and sometimes involve direct payments from originator companies to generic companies. In total, these payments amounted to more than € 200 million
4. *Interventions before regulatory bodies*, which have to approve generic products and decide on their pricing and reimbursement status. These interventions slow down the approval process

by 4 months on average.

Where successful, these practices result in significant additional costs for public health budgets—and ultimately consumers—and reduce incentives to innovate.

That was entirely nonsense:

“Patent Clustering.” Of course there is patent clustering—there always has been. Some now speak of “thickets” and postulate that there are:

*a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology.*¹⁰

or

*an overlapping set of patent rights which require innovators to reach licensing deals for multiple patents from multiple sources.*¹¹

This is said to lead to “hold-up”—economists are fond of emotive tags. But there is no evidence of any such thing. Just look at these innovative industries and ask, if things are moving so fast, “where is this ‘hold-up?’” In the real world, clusters of patents are, and always have been, a sign of a technology in a state of high inventiveness and rapid change—the opposite of hold up. Edison surrounded himself with a thousand patents, the sewing machine wars of the 1850s and ‘60s were replete with patents,¹² and even Boulton, the business genius who, with Watt, was a key figure of the industrial revolution, used the patent system with clusters.

“A large number of litigation cases.” This was simply wrong: most pharma patents, like most patents in general, are not litigated. But that some important ones are is hardly surprising. There is a lot of money involved. Generic companies naturally challenge big pharma patents if there is a realistic possibility of winning. Equally naturally big pharma sues on its patents if there is a realistic possibility that they may be held valid. This sort of patent fight is and always has been a normal part of the pharma industry.

“Patent settlements restraining generic entry” is uncommercial and unrealistic. The Commission is worried about is a deal between a patentee and a generic company by which they settle a dispute about the validity of a patent by agreeing on a date, somewhere between the date of the agreement and the date of expiry, upon which the generic may enter the market. Money may pass. The emotive tag is “pay for delay.” The theory is that is anticompetitive because the generic agrees not to enter the market earlier than the agreed date. But the theory ignores reality for three distinct reasons.

First, put aside the rare case where the patent is surely invalid and the patentee knows that.¹³ The normal case is where there is real uncertainty about validity. Neither side knows who would win the prospective legal battle which if it takes place, will be expensive and time consuming. If the patent is valid then the agreement, far from being anti-competitive, is pro-competitive—it allows entry earlier than expiry. If the patent is invalid, then the agreement delays entry, but just by that generic company. However at the time of the agreement, no one knows whether it is or is not valid. That could only be determined by the end of final appeals in the very battle the agreement is designed to avoid. At the very least the Commission, if it wants to prove that such an

agreement is in fact anticompetitive, must surely prove the patent invalid.

The second reason is the real danger of deterring bona fide settlements. If these agreements are treated *ipso facto* as anticompetitive there is only one alternative: litigation to the death. The Commission has never really been able to cope with the commercial settlement of IP disputes where one party agrees not to do something.¹⁴ It is time it understood that settlement of IP disputes often involves one party agreeing not to do something. If that settlement is genuine and not bogus the Commission should keep out. If the Commission wants to prove that the agreement is bogus—a sham to cover what is, in reality, a market sharing agreement—then the onus must lie on it to prove it.

The third reason why this sort of agreement is not anticompetitive is that it in no way affects the right of any other generic to challenge the patent and to try to enter the market meanwhile. An agreement with just one generic cannot seriously affect the potential market in the drug concerned at all. You simply can't "pay for delay" by a deal with one company if anyone else who wants to sell the generic medicine is free to challenge the patent and try to enter the market meanwhile (subject to an interim injunction¹⁵). The position in Europe is not quite the same as in the US, where a first generic may in some cases get a limited period of exclusivity (see below) and cannot "clear the way" by attacking the patent well before intended marketing.

"Interventions in the courts with decisions of regulatory bodies" is equally fallacious. The report itself concedes that about a third of these interventions, are justified. Moreover, the legality of actions of administrative authorities must be open to challenge. Just because pharma companies sometimes challenge unsuccessfully is hardly a reason for saying they are doing anything wrong. DG Comp is saying that the very act of going to court is wrong—a point to which I return.

IT IS TIME IT UNDERSTOOD THAT SETTLEMENT OF IP DISPUTES OFTEN INVOLVES ONE PARTY AGREEING NOT TO DO SOMETHING. IF THAT SETTLEMENT IS GENUINE AND NOT BOGUS THE COMMISSION SHOULD KEEP OUT. IF THE COMMISSION WANTS TO PROVE THAT THE AGREEMENT IS BOGUS—A SHAM TO COVER WHAT IS, IN REALITY, A MARKET SHARING AGREEMENT—THEN THE ONUS MUST LIE ON IT TO PROVE IT.

In the end, the Sector Inquiry largely fizzled out but not before a very large amount of public and pharma company money was wasted. The Commission has, to its undemocratic shame, never said how much it cost the public purse. It is difficult to imagine that it involved less than 30 expensive officials for at least two years; a better estimate might be 50.¹⁶ As for the industry costs, even after the initial raids, each company was swamped with ill-thought-out questions on a weekly or less basis. It must have cost the industry as a whole of the order of half billion euros. A scandal that has not had the publicity it deserved.

All that is left is "monitoring" of pharma settlements. At first this seemed to be largely to save face. For initially there was only the investigation (still ongoing) of the *Servier* case where there indeed may be a case to answer.¹⁷ It is possible that it knew full well that the patent it sought to enforce was invalid.¹⁸ But this year, probably encouraged by US developments, DG Comp has gone on the warpath against pay for delay settlements. It may that some cases of this are justified—those where no patent validity settlement is involved.¹⁹ But others are very worrying indeed; for instance, the huge fine of €93 million on Lundbeck for settling patent litigation

with a number of generic companies.²⁰ I am glad to see that Lundbeck are fighting this.²¹ In a rational world they should win.

The final Pharma report was accompanied by a statement from then Competition Commissioner Neelie Kroes, who, unveiling the findings of an 18-month inquiry into the pharmaceuticals sector, said:

“There is something rotten in the state [of the pharmaceutical industry]. Makers of original medicines are actively trying to delay the entry of generic medicines on to their markets.” (18th July 2009).

Others may think there is something ill—a wrong mindset—within the Competition Directorate revealed by this. It was a wholly unjustified slur on a whole innovative industry: an industry which acted in a perfectly rational and legally justified way; an industry which spends (with great risk) 17 percent of its *income* on trying to find new or better medicines for humanity. And which, unlike lawyers, economists, judges, or officials, actually prevents people from dying or suffering.

Nowhere in its inquiry did the Commission look to see what profits were being made by the grasshoppers—the generic companies whose very business depends on the earlier inventive work of the ants, still less of how much or little of those profits were being spent on research.

I have just one thing to add by way of postscript in relation to this sorry story. As I have said, the Commission clearly knew almost nothing about how the patent system worked. This is hardly surprising—it has no staff of its own experienced in this. True it is that it sought some assistance from the EPO who sent a very able patent examiner to assist. But patent examiners are not the people to ask about the commercial working of the system after patents are granted. Their experience is about examining for validity, not the commercial exploitation or enforcement of patents. Once a patent is granted, it goes out into the wide commercial world, leaving the Office which gave birth to it in ignorance of its adult life.

The FTC and Pharma

The competition authorities in the USA, particularly the Federal Trade Commission, have been equally aggressive to innovative pharma in the USA. Although nothing like the full-blooded assault on a whole industry by DG Comp. is possible in the USA, the FTC’s campaign—almost vendetta²² against pharma—has been and is intensive. The key case is *FTC v Actavis*.²³

Before I discuss the US further, a difference between generic entry in the US and in Europe should be noted. In the US, the first generic is given a special position by the Hatch-Waxman Act. It is not only allowed to enter the market with regulatory authority without having to undergo a duplicative application process (Europe is in broadly the same position), but in addition, in some circumstances, it may get a 180-day monopoly as the first generic allowed on the market. This has to be before expiry. Europe has no equivalent; any generic with regulatory approval can come on the market once the patent is out of the way by expiry or earlier revocation. In Europe, the first generic gets no legal advantage from being first.

There is another significant difference between Europe and the USA. In Europe, there is no “standing”

requirement for a party who wishes to ask a court to revoke a patent. Anyone can challenge at any time. Any generic feeling itself impaired by a patent can attack it well ahead of its intended marketing. It is for that reason that the UK has a “clear the way” rule about interim injunctions. A pharma patent holder will normally get such an injunction against a generic company about to come on to the market but who failed to attack the patent earlier.²⁴ In the US there is a “standing” requirement before a party can seek a declaration of invalidity:²⁵ a generic company cannot, well before it intends to market, ask a court to clear the way.

I turn to *Actavis*. Solvay (the patentee) made a reverse payment to Actavis in return for Actavis agreeing not to enter the market for a set term and settling the ongoing patent litigation the key issue of which was the counterclaim for invalidity. The FTC said this was a violation of antitrust laws. It took a brutal, simplistic per-se anti-patentee stance:

“A payment from one business to another in exchange for the recipient’s agreement not to compete is a paradigmatic antitrust trust violation. The question presented here is whether such a payment should be treated as lawful when it is encompassed within the settlement of a patent infringement suit. The answer to that question is no.”²⁶

Throughout the argument there was simply no discussion of the effect sought by the FTC—less income for pharma companies with the inevitable depressing effect on research funding.

The Court came up with an unsatisfactory result: in effect, “that it all depends.” It rejected the FTC black and white position that all pay for delay settlements are inherently bad and rejected the alternative view that provided the settlement was of a bona fide dispute; antitrust law had no place. It said that each case should be judged by a “rule of reason.” But no-one knows what that means, with the result that the case (and others) is ongoing with much uncertainty. There is much to be said for the dissenting opinion of Roberts CJ to which two other members of the court assented.

For present purposes there are two things to note: first is that the FTC intervention has made it much more difficult to settle *bona fide* patent litigation between big pharma and generics. And secondly it was essentially anti-patent; the sooner patent protections falls away, the better was the driver behind the FTC’s approach. The result is less protection for the innovator companies who will have less income to fund their research. It would have been much better for our future medicines if the FTC had let well alone. The same thinking has now spread to Europe. I hope it is rejected.

Patents and Telecoms: Standard Essential Patents (“SEPs”)

I turn to the other area where the regulatory authorities’ interventions are mistaken and dangerous for innovation. It concerns “standard essential patents,” or “SEPs.” Many industries find it necessary to develop technical “standards” so that the products of one company can work with those of others. Standards are very old—standards for things like telephones go back to the 19th century, and for gramophone records not much later. Today we have many more standards—for broadcasting, Blu-ray, CDs, DVDs and on. Most prominent are the standards for mobile phones. Over the years, inventors have allowed their inventions to be used for standards without any problem. The general nature of the system is that all parties allow use of their patented inventions, either via licensing or a patent pool.

In particular, for mobile telephony, the standards are set by an industry-wide organization called ETSI.²⁷ It sets the standards for 3G, 4G and the future 5G, 6G, and so on. ETSI members get together to settle on the standard. It is a hugely complicated process requiring many thousands of engineers' hours. The aim is to make the standard work as well as possible—the better it does the more the market will want the parties' products.

DG Comp suspects that manufacturers vie to get their particular patented solution made part of the standard with a view to a large income stream later. It drew this inference from a study which showed a spike in patent applications shortly before a standard setting meeting. Again, this shows a lack of understanding of the patent system—you have to apply for a patent before you disclose your invention (as you will at ETSI) or your patent will lack novelty. It makes entire sense to apply for your patent just before you disclose because then you can put the most information about your invention into your application—also important for validity (insufficient disclosure being a ground of invalidity).²⁸ There is no evidence to support the Commission's hypothesis—the competition is between qualities of solutions to get the best standard—the best standard is the motivation.

THE COMMISSION TAKES THE VIEW THAT THE VERY ACT OF ASKING FOR AN INJUNCTION IN HIS CLAIM IS AN ABUSE OF MONOPOLY FOR WHICH THE PATENTEE CAN BE FINED AND ENJOINED.... ONE CAN HARDLY BELIEVE HOW WRONG THIS IS...

ETSI members must make a FRAND (Fair Reasonable and Non-Discriminatory) commitment to ETSI—to offer licences under their patents for use with the standard on Fair Reasonable and Non-Discriminatory terms. Other standard bodies sometimes use the acronym RAND.²⁹ The commitments are not the same for all standards.³⁰

No-one checks whether declared patents really are essential—or even reads the patents or patent applications. You could even declare a patent on a rubber boot as essential and it would be recorded as such. Broadly, over-declaration is a good thing—much better to err on the side of that than on the side of under-declaration with the result that an undeclared patent might cover the standard.³¹

The wording of the commitment to ETSI is that the patentee will give:

“...an irrevocable undertaking in writing that it is prepared to grant irrevocable licences on fair, reasonable and non-discriminatory terms and conditions under such IPR to at least the following extent [then follow manufacture and use details].”

Note that it is the patentee who commits to grant a licence on FRAND terms. It follows that if he makes an offer which is FRAND compliant he has complied with his commitment. The licensee has but to accept and the grant is made. The position of the would-be licensee, whether or not he is a “willing licensee” in some sort of loose sense or is prepared to enter into negotiations, is quite irrelevant.

So what happens if a SEP patentee sues someone on his patent? The Commission takes the view that the very act of asking for an injunction in his claim is an abuse of monopoly for which the patentee can be fined and enjoined. Here is the Press Release in *Motorola v Apple*:³²

The European Commission has informed Motorola Mobility of its preliminary view that the company's seeking and enforcing of an injunction against Apple in Germany on the basis of its mobile phone standard-essential patents ("SEPs") amounts to an abuse of a dominant position prohibited by EU antitrust rules. While recourse to injunctions is a possible remedy for patent infringements, such conduct may be abusive where SEPs are concerned and the potential licensee is willing to enter into a licence on Fair, Reasonable and Non-Discriminatory (so-called "FRAND") terms. In such a situation, the Commission considers at this stage that dominant SEP holders should not have recourse to injunctions, which generally involve a prohibition to sell the product infringing the patent, in order to distort licensing negotiations and impose unjustified licensing terms on patent licensees. Such misuse of SEPs could ultimately harm consumers.³³

One can hardly believe how wrong this is:

1. It is saying that merely going to court to ask for an order is an abuse of monopoly. That breaches two important principles:

(a) First at a very high level it breaches the right of a party's access to the courts contained in Art. 6 of the ECHR and well recognised in case such as *Golder v. UK*.³⁴ The Commission is in effect standing with a shotgun at the courthouse door and saying "if you go in there and dare even ask for an injunction, we will shoot you." Contrast that with the European Court of Human Rights in *Golder*:

The principle whereby a civil claim must be capable of being submitted to a judge ranks as one of the universally "recognised" fundamental principles of law; the same is true of the principle of international law which forbids the denial of justice. Article 6(1) must be read in the light of these principles. Were Article 6(1) to be understood as concerning exclusively the conduct of an action which had already been initiated before a court, a Contracting State could, without acting in breach of that text, do away with its courts, or take away their jurisdiction to determine certain classes of civil actions and entrust it to organs dependent on the Government. Such assumptions, indissociable from a danger of arbitrary power, would have serious consequences which are repugnant to the aforementioned principles and which the Court cannot overlook.

(b) Secondly, still at a high level, it breaches the principle of sincere co-operation in Art 4(3) of the bedrock Treaty on European Union. This provides:

Pursuant to the principle of sincere cooperation, the Union and the Member States shall, in full mutual respect, assist each other in carrying out tasks which flow from the Treaties.

The Commission is really saying, "we do not trust the Member States' courts—they might grant an injunction which we think would be an abuse of monopoly. We, an administrative agency, know better." So much for "mutual respect."

2. It is on the facts perfectly ridiculous—to suppose you can bully Apple, ZTE or Hauwei³⁵ by a mere application to court shows a profound ignorance of how real big businesses or courts work. If these companies

have a defence of abuse of monopoly they will surely deploy it with maximum force. None need the Commission to defend them.

3. It is virtually a *per se* rule—for although there is some concession in the case of a defendant who might be impecunious, it is difficult to see how that could work in practice. Suppose the patentee considered the defendant impecunious but the Commission thought otherwise? That the Commission takes the part of such big companies itself demonstrates the essentially *per se* nature of its position.

4. It overlooks the fact that before the Court decides whether or not to grant an injunction, it will not only hear the parties but could hear the Commission, too. For the Commission has a right to intervene.³⁶ Surely the Commission should take the least intrusive course of intervening rather than threatening fines?

5. It ignores the fact that litigation is a way of bringing things to a head. The pressure of a date in court is not a pressure to do a non-FRAND deal, but a pressure to do a FRAND deal. All defendants string things out if they can. Litigation is a continuation of negotiation by other means.

6. The Commission's focus on a "willing licensee" is entirely misplaced. Defendants will always say they are willing to negotiate—that really means stringing things out for as long as possible and paying as little as possible in the end, a sensible commercial tactic if you can get away with it. There may, of course, be negotiations, but all that matters legally is whether the patentee has made a FRAND offer. The focus should be on that, not what the defendant says is FRAND or whether he says he is willing to negotiate. The concept of a "willing licensee" is amorphous and impossible to pin down.

7. It asks the question of "abuse" at the wrong time—before the court is asked to make its decision. It treats an application for an injunction as if it were an injunction. When the court comes to make its decision it will have evidence from both sides and can decide what to do.³⁷

8. It ignores the fact that if the patentee has made no offer *when the court makes its decision* the court will surely refuse an injunction because the patentee has committed to make a FRAND offer—a commitment which the industry largely accepts is binding on all, never mind any legal theory which might suggest otherwise.³⁸

9. It assumes that a declared SEP is in itself market dominant. But a declared SEP may not be essential at all—many are not. Even an SEP, once essential, may cease to be so because ways around have been devised.

10. If the patentee has made an offer, it can be examined to see whether it is FRAND or not. That can be determined by the court if necessary, but there may be better ways, such as arbitration by an expert panel. Meanwhile, the court may make an interim order requiring the defendant to pay a minimum to the patentee, and the disputed difference into escrow. A solution I favour.

11. It ignores the reality that if the patentee is deprived of any right to an injunction or possibility of one, defendants will simply fight infringement and validity patent by patent. If they lose, why then all they have to pay is damages to be assessed on FRAND terms—"reverse hold-up" with a vengeance. The Commission's actions support just that. How much have Apple, Huawei and ZTE paid anyone for use of the standards on which they have been able to build their businesses?

12. It ignores the reality that this industry is one of the most competitive in the world—which company will be leader even in 3 years’ time? The competition is in products and innovation. Where is there a competition law problem?

13. None of this is in the consumer’s interest. If the non-innovators have to pay only little and late, does anyone think it will be passed on to the public? No, it will mean that the profits of the grasshopper companies will be greater at the expense of the ant innovators, who made the grasshoppers’ business possible. And consumers will lose by reason of less future innovation.

I suspect the Commission knows it is on shaky ground. For it is pursuing an insidious course of negotiating with patentees to get them to agree to limit their rights to go to court. It recently³⁹ issued a press notice:

The Commission has concerns that Samsung’s seeking of injunctions against Apple in the European Economic Area (EEA) on the basis of its mobile SEPs may have amounted to an abuse of a dominant position prohibited by EU antitrust rules (see IP/12/1448). To remedy these concerns, Samsung has offered to abstain from seeking injunctions for mobile SEPs for a period of five years against any company that agrees to a particular licensing framework. Interested parties can now submit their comments within one month. If the Commission concludes, in light of the comments received, that the commitments address the competition concerns, it may decide to make them legally binding on Samsung.

Seeking injunctions before courts is generally a legitimate remedy for patent holders in case of patent infringements. However, access to patents which are standard-essential is a precondition for any company to sell interoperable products in the market.

The Commission considers that the seeking of an injunction based on SEPs may constitute an abuse of a dominant position if a SEP holder has given a commitment to license its SEPs on Fair Reasonable And Non-Discriminatory (FRAND) terms and where the company against which an injunction is sought is willing to enter into a licence agreement on FRAND terms.

The Commission is concerned that the seeking of injunctions in such circumstances could allow Samsung to impose royalty rates or other licensing terms, such as broad cross-licenses, which a licensee would not agree to, absent the threat of having its products excluded from the market. This may unduly distort licensing negotiations and cause harm to consumers by increasing prices, reducing product choice and stifling differentiating innovation in the markets for smartphones and tablets.

This is trying to create precedent by creep. Undertakings of this sort imply that they are necessary. They are well attacked by Prof. Marsden’s, “Soft Law. The Emperor’s Clothes Laid Bare: Commitments Creating the Appearance of Law while Denying Access to Law.”⁴⁰

The Commission aims to extract the undertaking by the threat of fines. Companies are really frightened of the Commission:⁴¹ it behoves it to use its powers with care. It did not have to threaten fines. It could simply have said to Samsung and Motorola, “We think you are wrong, let us go to the CJEU and test the legality of your actions. We will levy fines only if we win and you carry on regardless.”

The time has come to stand up—not to pay Danegeld.⁴² Companies should say to the Commission: “See you in court. You can’t use your powers to deny access to the courts. You can’t allege abuse of monopoly, hold up and the like based on a theory which unsubstantiated on real facts but is obviously wrong given the rate of innovation in this sector. You should not be protecting those who contributed virtually nothing from the inventions which make up the standards. Leave it to the Courts; they can enforce the FRAND obligations.”

One other matter: the Commission’s current stance may have an unintended consequence—that it becomes a FRAND determining tribunal. On 21st November 2013, it was announced that Sierra Wireless had complained to the Commission that Nokia would not grant it a FRAND licence. What if Nokia’s response is that it had made a FRAND offer? The Commission will have to decide one way or the other. Surely the Commission is not the right tribunal for determining whether an offer is or is not FRAND. Yet it could find itself in a mire of complaints and counter-complaints.

All of the above applies in the US, *mutatis mutandis*. The FTC and Justice Department successfully persuaded President Obama to veto the import ban on Apple products which the ITC, following a reasoned decision of its administrative court, held infringed a valid patent of Samsung.⁴³ The theory was that the patent was subject to

THE COMPETITION AUTHORITIES’
ANTI-PATENTEE ACTIONS PENALISE
THOSE WHO MADE THE INVENTIONS
IN THE PAST... THE COMPETITION
AUTHORITIES SHOULD CEASE
HARASSING INVENTIVE INDUSTRIES,
REMEMBER THAT PATENTS EXPIRE
ANYWAY AND LET THE PATENT
SYSTEM DO THE JOB IT WAS DESIGNED
TO DO. LEAVE THE ANTS ALONE.

a FRAND declaration. The court had not accepted the FRAND defence. To an outsider it looks awfully like a Government overruling a court decision—something rather incompatible with the rule of law.

I end with this. The Competition Authorities’ anti-patentee actions penalise those who made the inventions in the past. Economists call the costs of making and developing those inventions (and all the abortive research and development) “sunk”—money already spent. If you say sunk costs produce results of little value today, you degrade the incentive to sink costs today—to do R&D for the inventions of tomorrow. For

inventions which might become the subject of SEPs, you are threatening not only invention but the very process of standardisation. If those who create the inventions for new and faster standards see the prospect of proper reward degraded then the rate of progress to 5G, 6G, and so on will diminish. Such is the public demand for more and faster capacity; the result could well be that existing standards will not cope—so that either our phone connections will clog up or the price of connection will have to go up to reduce demand. And if the pharma companies have less income, we will have less new medicines in the future.

The Competition Authorities should cease harassing inventive industries, remember that patents expire anyway and let the patent system do the job it was designed to do. Leave the ants alone. ▲

1. Hugh Laddie Professor of Intellectual Property, University College London, a former Lord Justice of Appeal of the Court of Appeal of England and Wales. I am grateful to Martin Adelman for trying to put me right about a couple of things. If I failed it is not his fault.

2. See e.g. my judicial criticism of the ECJ's decision that a trader in cheap perfume cannot honestly say his perfume smells like a famous brand, *L'Oréal v Bellure (No.2)* [2010] EWCA Civ 535.
3. 21 if you use the priority system well.
4. Indeed it is unfortunate that patent law does not go further in some cases – particularly in not really providing adequate protection for future uses of known medicines. This is a subject of towering importance. One major pharma company says it is as important as the search for new medicines. Some generic companies are of the same view: they too would like to do research for new uses for medicines they already make. Governments and the Commission would do well to think how such research can be incentivised.
5. Not always, for instance Holland abolished patents from 1869 to 1912. Maybe that enabled a backward economy to catch up. Once it did patents became vital to future progress. Philips, which started in that patent-free period, is now one of the major innovative (and hence patenting) companies of the world.
6. Or what they believe will be lower prices. Often their grasshopper protégées will simply keep the profit.
7. Intellectual Property Rights Intensive Industries Contribution to Economic Performance and Employment in the EU,
8. I quoted it the public debate about the preliminary findings of the Commission's Sector Inquiry into the pharma industry, 28th November 2008.
9. 8th July 2009. Much of the nonsense in the Preliminary Report was dropped.
10. C Shapiro, 'Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting' in AB Jaffe et al (eds), *Innovation Policy and the Economy* (MIT Press, 2001).
11. *Hargreaves Report* (n 8) 18.
12. See Mossoff: The Rise and Fall of the First American Patent Thicket: The Sewing Machine War of the 1850s *Arizona Law Review*, Vol. 53:165.
13. Such cases are very rare – it may well be that the law ought to have a remedy to deal with such case – in some countries it indeed may, under some sort of tort of *abus de droit*.
14. 14 In the 1970s I was involved in a Commission attack on a bona fide settlement of a trade mark dispute where the defendant agreed not only to change its name but not to use the plaintiff's name.
15. 15 Which can only be obtained on the terms that the patentee must compensate the defendant if in the end the patent is found invalid.
16. One pharma company told me that its dawn raid party had 7 officials.
17. The law may have something to say about an IP right dishonestly obtained. , but it is not clear that it is competition law: how can there be an abuse of monopoly when the complaint is that there was no monopoly only a purported one? The ECJ did not consider this point in *AstraZenica v EC Commission Case C-457/10P*. As far as the court was concerned the finding of abuse of monopoly turned on the obtaining of an IP right by a deliberate misrepresentation (see para. [99]). I have no real problem with the result. Liars deserve what they get. For present purposes the point to note is that the Commission apparently contended for something much more far-reaching, namely abuse of monopoly if the right granted was made by any misrepresentation, even if not dishonest – such is the tenor of the Commission decision of 15th June 2005.
18. See my Judgment in *Les Laboratoires Servier v Apotex* [2008] EWCA Civ 445 “This is the sort of patent which gives the patent system a bad name.”
19. As may be the case of *J&J* and *Novartis/Sandoz* see IP/13/1233 10th December 2013.
20. 19th June 2013. With a politician's bombast, Commissioner Almunia said: “Agreements of this type directly harm patients and national health systems which are already under tight budgetary restraints.” He is quite wrong if the patents are valid – pre-expiry entry then has a price reducing effect. And the Commissioner failed to mention the effect on future R&D – the fine alone will surely affect this.
21. Statement of September 2nd 2013

22. The Chairwoman of the FTC described the FTC's current challenges to patent settlements as a "mission", Law60, November 2013.
23. 570 U.S. ____ (2013)
24. *Novartis v Hospira UK* [2013] EWCA Civ 583.
25. The US does not actually have a formal revocation proceeding. "Standing" requires that the patentee has in some way indicated that it will try to enforce the patent. Merely owning it is not enough. Makes no sense to me: a patent is a running public claim to a monopoly, the public should be able to attack it at any time.
26. Deputy Solicitor General Malcolm Stewart opening the oral argument on March 25th 2013. A little later he qualified it by saying such an agreement was "presumptively unlawful" but that the presumption could be rebutted (he did not say within any precision how.)
27. The European Telecommunications Standards Institute.
28. I suspect another influence too – it is human nature to do things at the last minute!
29. No-one suggests that "Fair" adds anything.
30. And it may not be a mere difference in wording: there may be a real difference between a commitment actually to grant a licence on RAND or FRAND terms, and a commitment to offer a licence on such terms. The difference is that a commitment to grant leaves it open as to who is to determine the terms. The commitment to offer FRAND leaves it to the *patentee* to determine the terms of his offer – anything within the range of FRAND will comply with his commitment.
31. However some overdeclarations are deliberately done: to some extent numbers of patents count in licensing negotiations. The English Courts have held they have jurisdiction over whether or not a patent declared to be is in fact essential, *Nokia v Interdigital* [2006] EWCA Civ 1618. It involved a severe case of overdeclaration, 29 patents had been declared essential but only one found to be so, *Nokia v Interdigital* [2007] EWHC 3077 (Pat). Validity was never considered.
32. 6th May 2013
33. The Press Release in the case of *Samsung v Apple*, 31st January 2012, is similarly worded though there is no reference to "enforcing" for the obvious reason that Samsung never got as far as a hearing in which an injunction was asked for.
34. Series A No 18, (1979-80) 1 E.H.R.R. 524. See also Art. 47 of the EU Charter of Human Rights, Case C-279/09 *DEB Deutsche Energiehandels etc. v Germany* [2010] ECR I-13849. There is plenty more to like effect. True it is that the right of access is not entirely absolute, but it is only in "wholly exceptional circumstances that "the fact that legal proceedings are brought is capable of constituting an abuse of an dominant position within the meaning of Article 86 of the Treaty, Case T-119/09 *Protégé International Ltd v Commission* (13 September 2012) at §§48-49. See also *ITT Promedia v Commission* [1998] ECR II-2937.
35. These are all latecomers to mobile telephony and have no or hardly any SEPS themselves. Moreover it seems they are not contributing to 4G, 5G and so on: all published figures show they have very few SEPs between them.
36. Art. 335 of the TFEU.
37. There is no suggestion of ex parte injunctions in this field.
38. There is some debate about whether the FRAND commitment given to ETSI is enforceable by a third party. In the US Judge Robert held the RAND commitment was. In Europe the legal position fairly clearly is that a third party can indeed require the patentee to make a FRAND offer. For ETSI is in France and the commitment is fairly obviously given under French law which admits of what is sometimes called the *jus tertius* – the right of a third party to enforce a contract made by others for its benefit. It seems German Court said the issue in Germany was governed by German law because the patent was German. That cannot be right: the FRAND commitment must be governed by a single law – the parties cannot have intended the chaos of multiple laws, some of which have the *jus tertius* and others not.
39. 17th October 2013
40. CPT Antitrust Chronicle, October 2013(1)
41. Not surprisingly. I heard that one pharma company which

42. See Kipling's poem of that name.
43. Decision of 3rd August 2013.

Intellectual Property Experimentalism By Way Of Competition Law

BY TIM WU¹

Competition law and Intellectual Property have divergent intellectual cultures—the former more pragmatic and experimentalist; the latter influenced by natural law and vested rights. The US Supreme Court decision in *Federal Trade Commission v. Actavis* is an intellectual victory for the former approach, one that suggests that antitrust law can and should be used to introduce greater scrutiny of the specific consequences of intellectual property grants.

One day at the Athenaeum club in London, an unnamed attorney said something along these lines:

“We know that the laws only penalize agreements that restrain competition. Given the existence of a valid patent, the market belongs to the drug manufacturer, and there is, therefore, no possibility of restraining competition.”

“You’re saying generic drug makers and pharmaceutical companies aren’t competitors,” said someone, with an air of disbelief.

“Of course not,” said the attorney, “because as an owner of a patent, they have been given the market, so there is no competition.”

“I can’t sit here and listen to this,” said someone else.

THE COMPETITION LAW IS MORE PRAGMATIC IN ORIENTATION, AND, BORROWING FROM THE NORMS OF ECONOMISTS, STRONGLY INTERESTED IN THE CONSEQUENCES OF THE LAW’S OPERATION. MEANWHILE THE INTELLECTUAL PROPERTY LAWS RETAIN A DOMINANT CULTURE CLOSER TO A NATURAL LAW RIGHTS TRADITION, TAKING THE PRIMARY DUTY OF LAW AS THE CREATION AND PROTECTION OF VESTED RIGHTS, WITH LESS REGARD FOR THE SPECIFIC EFFECTS OF RIGHTS GRANTED.

The exchange of words above, one of many heated exchanges during that day, may seem like nothing more than a typical clash between those who happen to take different views of a case. But disagreements can sometimes reveal something deeper, and what comes across here is a fundamental intellectual divide between practitioners in the intellectual property and competition law fields. While sometimes said to share “a common purpose of promoting innovation and consumer welfare,”² those with any experience in the two fields know how different the thinking in each can be. The competition law is more pragmatic in orientation, and, borrowing from the norms of economists, strongly interested in the consequences of the law’s operation. Meanwhile the intellectual property laws retain a dominant culture closer to a natural law rights

tradition, taking the primary duty of law as the creation and protection of vested rights, with less regard for the specific effects of rights granted.

While I'm hardly the first to write about the distinctive intellectual traditions of the two regimes,³ in this paper I want to try to take a new look at the intellectual divide, and in particular, how it yields different approaches to a key issue, namely, how the law reacts to changing conditions and unintended consequences.

I am not neutral party in this debate. I believe that the IP laws in general, and the patent laws in particular, need to be better attuned to their actual effects on commerce. Stated differently, I think we need a more *experimentalist* intellectual property regime, that is, one that views the laws as an ongoing project in constant need of improvement.⁴ As a practical matter, however, such regime change seems unlikely to emerge organically. One means for introducing a more error-sensitive intellectual property system is increased use of the competition laws as an oversight regime.

That's a task that competition enforcers have already been taking seriously, and the Supreme Court's 2013 decision in *Federal Trade Commission v. Actavis* can be read (perhaps generously) as a blessing of such efforts. The decision effectively approved more than a decade of efforts by the Federal Trade Commission to stop "pay-for-delay" patent settlements that it considered anticompetitive. At a more general level, it approved of an ongoing examination of the actual consequences of patents using the antitrust laws, and as such the details of the decision are less important than its approach. Of course, implicit in the grant of a patent is accepting anticompetitive consequences, but in *Actavis* the Court viewed that fact merely as another factor for antitrust analysis. In other words, one job of the competition laws is improving the patent system and perhaps intellectual property more generally.

I. EXPERIMENTALISM

Experimentalism is not a word that attorneys use very often. At its most general, the idea of legal experimentalism is to apply the scientific method of hypothesis, experiment, and observation of consequence to challenging legal and policy problems. It is, as such, closely related to a "pragmatist" legal philosophy.⁵

John Dewey is usually credited with laying a philosophical foundation for policy experimentalism in his writings in the 1910s and 1920s. Dewey, whose background was in education, believed that a successful democracy needed the capacity to learn and improve itself. The key to learning, he believed, was the processing of experiences, or in his words the "reconstruction or reorganization of experience which adds to the meaning of experience and which increases ability to direct the course of subsequent experience."⁶

As relevant to the legal system, Dewey thought policy and "proposals for social action" should be subject to the experimental method. Policy-making, he said, should be a constant process of learning from experience, rather than relying on rigid or foundational truths. "Policies," Dewey argued, should be "experimental in the sense that they will be entertained subject to constant and well-equipped observation of the consequences they entail when acted upon, and subject to ready and flexible revision in the light of observed consequences."⁷

As understood here we can describe legal experimentalism as comprising three main principles. First, for the experimentalist, laws are simply instruments meant to achieve some end and useful only to the extent they do so. A law has no intrinsic value, and its existence should not necessarily count in favor of its retention.

Second, every law should be thought of as an ongoing experiment. That is to say, every enactment, regu-

lation or judicial opinion must be seen as that moment's best guess as to what a rule should be, in light of imperfect information and human fallibility. Borrowing Dewey's language, policies should be thought of as a "working hypothesis, not as programs to be rigidly adhered to and executed."⁸ Given the imperfect nature of law-making, policy should be subject to revision when faced with new information or changed conditions. The law must also be able to learn and improve itself based on observation of consequences, intended or otherwise.

Third, in service of the first two principles, the relevant institutions should be designed to learn and improve the law through an experimental process. There are, actually, many ways this could be achieved.⁹ But to actually perform something that resembles an experiment, its most important structure is some tolerance for legal diversity, coupled with a centralized mechanism for evaluating the results of different approaches to the same problem.

II. COMPETITION LAW

Over the last few decades, legal regimes that are experimentalist by design have emerged in discrete areas of regulation, such as mine safety or in the use of drug courts. As Professors Charles Sabel¹⁰ and Bill Simon put it, these are institutions that achieve a "decentralization of operative control with central coordination of the evaluation of results."¹¹ But it is also possible for legal experimentalism to arise organically, particularly when scattered lawmakers confront similar problems repeatedly. Arguably, for example, a form of serial experimentalism has been going on for centuries within the framework in the common law tradition, where diverse judges work through cases, continually reassessing old precedent and occasionally discarding rules that no longer work.¹²

ARGUABLY, FOR EXAMPLE, A FORM OF SERIAL EXPERIMENTALISM HAS BEEN GOING ON FOR CENTURIES WITHIN THE FRAMEWORK IN THE COMMON LAW TRADITION, WHERE DIVERSE JUDGES WORK THROUGH CASES, CONTINUALLY REASSESSING OLD PRECEDENT AND OCCASIONALLY DISCARDING RULES THAT NO LONGER WORK.¹² WHILE COMPETITION LAW ISN'T FORMALLY EXPERIMENTALIST, I SUGGEST THAT THE COMPETITION AND ANTITRUST LAWS TEND TOWARD SUCH INFORMAL EXPERIMENTALISM.

While competition law isn't formally experimentalist, I suggest that the competition and antitrust laws tend toward such informal experimentalism. It is, first, evident that the competition law has long been considered an instrument for achieving goals, not recognizing abstract rights. The law's goals are general to a fault: once upon a time in the United States, "to eliminate the evil of bigness;"¹³ in more recent decades, the less colorful goal of "encourage[ing] competitive markets to promote consumer welfare."¹⁴ Even if the goals of the law have evolved, the idea that the law is only really understood as valuable to the degree it promotes those goals has not.¹⁵ None of this is to say that antitrust attorneys never get fixed in their views, but that isn't usually seen as an attractive trait.

Structurally, over the last several decades, the geographic spread of competition laws has created a natural experimentalist structure. The exact same legal problem is often faced by the European authorities, the two American agencies (Department of Justice and the Federal Trade Commission), and hundreds other national competition authorities, American states, and individual American judges in private lawsuits. Furthermore, within the large agencies, individual staff attorneys often have a surprising discretion to begin informal investigations.

That experimentalist structure has been preserved, because, despite the obvious potential for international conflict, there is no serious movement toward a single, uniform competition law, and there is no global competition law treaty. As Hugh Hollman, Bill Kovacic and Andrew Robertson put it, “complete uniformity is probably unattainable and undesirable” for it would impede experimentation.¹⁶ As we shall see, that’s an attitude sharply at odds with the position taken in the intellectual property community.

Finally, for more than a century in the United States, and for decades in Europe and the rest of the world, the law has undergone ongoing tinkering and serial experimentation by legislatures, agencies, and courts. That is true both within regimes and across them. Consider, for example, the specific history of Section 2 of the Sherman Act, which has gone through various peaks and valleys not unlike a long-running cardiogram. Based on then-current opinion, the law has gone from being close to a no-fault deconcentration rule¹⁷ to something very close to dead (during the second Bush administration).¹⁸ Or consider that there have been six versions of the American merger guidelines since 1968.

THE GLOBAL COMPETITION SYSTEM ALSO LACKS ANY FORMAL SYSTEM FOR A CENTRALIZED ASSESSMENT OF DIFFERENT APPROACHES. HOWEVER THERE ARE INFORMAL MECHANISMS FOR DOING SO, INCLUDING THE INTERNATIONAL COMPETITION NETWORK, THE OECD WORKING GROUP AND OTHER LESS FORMAL GROUPS THAT SPEND THEIR TIME ASSESSING BEST PRACTICES AND, INFORMALLY, COMPARING RESULTS.

Europe has, if anything, seen greater variation. Consider, for example, the initial years of the European Commission Merger Regulation focused on the formal distinction between “cooperative” and “concentrative” joint ventures.¹⁹ European merger control has quickly evolved both in application and legislatively. Throughout the ‘90s the European Commission entertained novel antitrust theories and remedies,²⁰ and by 2004, just 15 years after the regulations were enacted, the ECMR saw substantial reform that, among other things, enabled the commission to prohibit mergers that “significantly impede effective competition” even if the merger would not confer dominance, which previously had been the sole standard.²¹

To be sure, this history of tinkering is not quite the formal, parallel experimentation that Sabel discusses. A critic might point out that Europeans don’t necessarily feel bound to learn from American failures, and vice versa. But what creates an experimentalist culture, as opposed to just legal diversity is a sense of continuing trying to learn from what has succeeded and what has failed. This may be due to the influence of economists than any other reason, and their fondness for the “natural experiment.”²² Of course, the success or failure of any law is hard and maybe impossible to measure accurately, which makes this a challenging proposition. Given the lack of controls and the difficulty of untangling cause and effect, it is difficult to know whether a given merger policy was more or less successful, or whether a major investigation was just a waste of resources or a critical turning point for an industry. But nonetheless there is a spirit of learning that is manifested in efforts to learn from failures, or imitate investigations seen as successful.

The global competition system also lacks any formal system for a centralized assessment of different approaches. However there are informal mechanisms for doing so, including the International Competition Network, the OECD working group and other less formal groups that spend their time assessing best practices and, informally, comparing results.

III. INTELLECTUAL PROPERTY LAWS

To say that the copyright, trademark and patent laws do not have an experimentalist spirit might be something of an understatement. Rather, to describe the laws as an ongoing experiment might easily risk offense.

There is no reason, *a priori*, that the intellectual property laws could not be subject to an experimentalist approach.²³ The laws have general goals and there are many potential ways to achieve them. To take a minor example, no one really knows what duration of copyright or patents is optimal. However, there seems at present to be very little apparent interest among lawmakers for trying different terms in different jurisdictions or for different subject matters and seeing what happens.

THE DOMINANT CULTURE OF INTELLECTUAL PROPERTY RETAINS A PERSISTENT LINKAGE WITH THE NATURAL LAW TRADITION, WHICH RECOGNIZES A NATURAL RIGHT IN THE INVENTOR OR CREATOR, AND THE LANGUAGE OF PROPERTY, OF WHICH THE PHRASE “INTELLECTUAL PROPERTY” IS JUST THE MOST OBVIOUS SIGN.²⁴

Even though the IP laws are sometimes discussed and defended in economic terms, one might say that the influence of economic thought has been partial. There is, to be sure, much discussion of incentives as the justification for intellectual property, but a full application of the scientific method has not generally been welcome. In fact, proposals by economists to tinker with obvious defects in the law are more often met with panic than interest.

Rather, despite the efforts of some scholars, the dominant culture of intellectual property retains a persistent linkage with the natural law tradition, which recognizes a natural right in the inventor or creator, and the language of property, of which the phrase “intellectual property” is just the most obvious sign.²⁴ In this view, the primary duty of the law is to recognize and protect rights of individuals (or corporate individuals). These could be rights against the government, like the American right against self-incrimination; rights in contract; or rights in property. The recognition of such rights is less a means to some other goal than the point of the law itself. What rights should be protected is, moreover, less a matter that depends on circumstance, but rather an answer that comes from consideration of fundamental truths.

The intellectual culture of the rights tradition has created two important differences with the competition regime in the face of changing conditions and the evolution of the law.

First, in the face of changing conditions, the logic of a vested right suggested that the priority should be the protection the right against erosion. Courts must defend the holder of a vested right not just against external threats, but against subsequent government action that might threaten the value of existing rights. Second, rights systems generally aspire toward uniformity across jurisdictions. Given a clear idea as to what rights should be protected, there is, logically, no clear reason for variation between jurisdictions or nations.

We can see the practical implications of both of these tendencies as they have manifested over the last few decades. For one thing, the reaction to technological change has usually been a fortification of the right rather than a rethinking of the system.²⁵ Consider, for example, the reaction of the laws to the massive technological changes over the 1980s and 1990s, when the successive development of personal computers, digitalization tech-

nologies and the Internet rather obviously changed the assumptions on which both patent and copyright had long relied. In both the United States and Europe the predominant legal response was to strengthen copyrights and to expand the subject matter of patent. The merits of such laws are not the subject here; the point was that, as opposed to reconsidering what would serve the broader purposes of the law under changed conditions, the challenge was framed as protecting the existing rights against potential erosion.

The goals of uniformity and predictability has had its clearest implications at the international level. Unlike competition law, which varies significantly between OECD nations, over the last several decades all of the IP laws have become subject to a much stronger and geographically broader web of harmonizing international agreements, on multinational, regional and bilateral levels. The general aim of these treaties is to homogenize the world's IP regimes, reducing or eliminating geographical variation. All of the major laws are the subject of longstanding global treaties specifying minimum protections (The Berne and Paris conventions), which were fortified in 1994 by the addition of an intellectual property agreement to the World Trade Organization, and further strengthened by numerous bilateral treaties since then. And of course the World Trade Organization, unlike the informal organizations common to competition law, has the power to punish deviations from the intellectual property treaties with serious trade sanctions.

The pattern can also be observed at the national level. Both in Europe and the United States the last few decades have witnessed many important measures taken to create uniformity. In the United States, a single appeals court, the Federal Circuit, has heard the nation's appeals in patent cases since 1982 in an effort to bring greater uniformity to the patent law. Though proposals for constructing a uniform patent court akin to the Federal Circuit in the European Union have been unsuccessful so far,²⁶ the European Patent Convention, founded in 1973, provides a common application for the prosecution of patents in each of the member states.²⁷

In short, stronger protection of uniform rights has been the clear trajectory of the intellectual property laws over the last few decades. That tendency is sharply at odds with the predispositions of the competition laws.

The dichotomy I am suggesting here is, of course, not absolute. In certain areas of the competition law, one can sense the influence of a vested rights theory, in, for example, the resistance to breakups of dominant firms, even if the economic case for doing so might be quite strong. And there are areas in IP law, like the American fair use doctrine (a judicial and scholarly favorite), which have, in fact, served as important outlets for judicial tinkering in the face of changing conditions. For example the famous *Sony* decision, blessing the VCR, broke with prevalent copyright doctrine, arguably as a reaction to perceived technological necessity.²⁸ Similarly, following a decade of bad press, Congress, the courts, and the American Patent Office have begun to make adjustments with American patent law. An example is the new post-grant review process, which includes a particular provision targeted at business method patents.

Nonetheless it would be hard to describe the intellectual culture of either the intellectual property laws as truly committed to experimental improvement of the law. It would be even harder to describe competition law as devoted to the protection of fundamental rights. We are left with a divergence in intellectual cultures with broad implications for just about every advanced economy in the world.

NONETHELESS IT WOULD BE HARD TO DESCRIBE THE INTELLECTUAL CULTURE OF EITHER THE INTELLECTUAL PROPERTY LAWS AS TRULY COMMITTED TO EXPERIMENTAL IMPROVEMENT OF THE LAW. IT WOULD BE EVEN HARDER TO DESCRIBE COMPETITION LAW AS DEVOTED TO THE PROTECTION OF FUNDAMENTAL RIGHTS. WE ARE LEFT WITH A DIVERGENCE IN INTELLECTUAL CULTURES WITH BROAD IMPLICATIONS FOR JUST ABOUT EVERY ADVANCED ECONOMY IN THE WORLD.

IV. USING ANTITRUST FOR PATENT EXPERIMENTALISM AT THE UNITED STATES SUPREME COURT

I believe there is a need for a more experimentalist approach to the intellectual property laws, and particularly to the patent laws. The law, I believe, needs better mechanisms not simply to celebrate its successes, but to correct its errors, both specific and general. One way this might be achieved is to act within the structure and institutions of the laws themselves; as just discussed, this is a project underway in certain respects. But the other path is to rely on the competition laws as a kind of oversight and adjustment mechanism for the intellectual property laws.

I am, of course, hardly the first to advocate this position, and it is clear that the world's major competition agencies already view part of their job as handling patents' unexpected consequences. Yet that role is not universally accepted for the ideological reasons just suggested. When a competition agency takes a look at how a patent or copyright is actually being used, it raises the question of what degree the government can, having given out rights, later declare that it has made a mistake and limit usage based on new evidence.²⁹ That was precisely the question presented in the American Supreme Court's 2013 patent-antitrust decision, *Federal Trade Commission v. Actavis, Inc.*, the Court's first case at the intersection of the two laws in quite some time.³⁰

At issue in the case was a particular practice of brand-name pharmaceutical drug manufacturers and their generic rivals with respect to patented drugs. In the late 1990s the Commission began investigating and challenging certain settlements between patent-holders and would-be generic entrants. The Commission suspected that generics and brand-name manufacturers were settling patent infringement cases in a way that effectively split patent profits at the expense of the public.

The facts of the particular case considered by the Supreme Court make clearer how this might happen. The case centered on AndroGel, a testosterone replacement therapy manufactured by Solvay, a relatively large drug company. In 2003, two other drug makers sought to introduce generic versions of the drug. While AndroGel is protected by a patent that expires in 2021, the generics challenged the patent as invalid, and also not infringed upon by the proposed generic products. Under the Hatch-Waxman Act, such challenges are a form of infringement, and Solvay sued.

Three years later the parties settled, agreeing to make generic AndroGel unavailable until 2015. Of course, the drug would have reached the public in 2006 or 2007 if the patent had been found invalid or not infringed. Meanwhile, Solvay made large payments to the generic manufacturers (which it claimed were unrelated). As such, the Federal Trade Commission alleged the settlement was a payment for a delayed arrival of a generic drug, or a "pay-for-delay" deal.

From the view of strict vested rights approach, it is obvious that such settlements ought not to be challengeable under the antitrust or other laws, for several reasons. The patent grant defines a right for a private citizen that is his with which to do whatever he wants. The idea of government examining, *ex poste*, the usage of rights already granted contradicts the idea of a right that has vested. Moreover, even if we might concede that the particular usage of the right in question is unattractive, the idea of selectively examining such rights *ex poste* threatens the certainty and clarity prized by the system as a whole.

This is what Chief Justice Roberts meant when he wrote, in dissent, that a patent right creates a “zone within which the patent holder may operate without facing antitrust liability.” Once assigned, in this vision the right is supreme within the zone defined by its claims. For the Chief Justice, so long as the patent holder had not left the scope of his assigned right, the case was over. As he put it, in a line that captures this notion, there was “no reason adjudicate questions of patent law under antitrust principles.”

To the Federal Trade Commission’s lawyers, its Commissioners and a majority of the Supreme Court, at least some of these so-called settlements were really just agreements to split the proceeds of a dubious monopoly at the consumer’s expense, and as such a specific failure of the patent system in need of correction. More broadly, since such settlements are not uncommon, they were, as a class, a defect in the system that should be fixed.

That framing found a receptive audience in Justice Breyer, the majority author, who among American judges is one of the most obvious in viewing the intellectual property law as an experiment, and a rather dubious one at that. In fact, Breyer’s previous writings represent some of the few pragmatist writings on intellectual property to emerge from the Supreme Court. He dissented in *Eldred v. Ashcroft*, a copyright case that examined a legislative effort to lengthen the copyright term, retroactively, in light of purportedly changing conditions.³¹ It is hard to see how a retroactive term extension promotes any new authorship, and so Breyer declared that the law’s “practical effect is not to promote, but to inhibit, the progress of Science.”

As a professor, Breyer was less restrained. In 1970 he wrote *The Uneasy Case for Copyright*, a lengthy piece that examined copyright and concluded that the law, if perhaps useful once, was no longer really necessary to ensure the production of creative works, and should therefore probably be done away with.³² Both of these works reveal a very different approach to legal change than that demanded by a rights-based model. The law is an instrument; we should examine its content in light of its claimed goals and, if current conditions suggest the law is valueless, so be it. But until 2013, Breyer had not yet had a chance to write at length on the patent-antitrust intersection.³³

The entire opinion is captured by one, early line: “[R]everse payment settlements can ... sometimes violate the antitrust laws.” The word “sometimes” is very dear to the competition law and the pragmatist/experimentalist method. It suggests uncertainty and a lack of fixed truths, that everything will turn on particularized assessment

THE ENTIRE OPINION IS CAPTURED BY ONE, EARLY LINE: “[R]EVERSE PAYMENT SETTLEMENTS CAN ... SOMETIMES VIOLATE THE ANTITRUST LAWS.” THE WORD “SOMETIMES” IS VERY DEAR TO THE COMPETITION LAW AND THE PRAGMATIST/EXPERIMENTALIST METHOD. IT SUGGESTS UNCERTAINTY AND A LACK OF FIXED TRUTHS, THAT EVERYTHING WILL TURN ON PARTICULARIZED ASSESSMENT OF FACTS AND CONSEQUENCES.

of facts and consequences. Breyer's opinion is at pains to emphasize the specific contingencies that could create anticompetitive consequences, as shown by the following passage with an astonishing string of contingencies (seven, by my count) that, in the right combination, may be more important than encouraging settlements.

[A] reverse payment, where large and unjustified, can bring with it the risk of significant anti-competitive effects; one who makes such a payment may be unable to explain and to justify it; such a firm or individual may well possess market power derived from the patent; a court, by examining the size of the payment, may well be able to assess its likely anticompetitive effects along with its potential justifications without litigating the validity of the patent; and parties may well find ways to settle patent disputes without the use of reverse payments.³⁴

Yet, by far the most significant thing about the opinion is not its use of semi-colons but rather the im-

PLICATION THAT THE ANTITRUST REGIME SITS IN A POSITION OF SUPREME OVERSIGHT OVER THE PATENT LAWS.³⁵ BREYER'S APPROACH MAKES CLEAR THAT, POTENTIALLY, ANY ANTICOMPETITIVE CONSEQUENCE THROWN OFF BY THE PATENT SYSTEM COULD BE SUBJECT TO CHALLENGE. TO BE SURE, A PATENT GIVES ITS OWNER SOME IMMUNITY TO ANTITRUST. BUT "PATENT AND ANTITRUST POLICIES ARE BOTH RELEVANT IN DETERMINING THE 'SCOPE OF THE PATENT MONOPOLY'—AND CONSEQUENTLY ANTITRUST LAW IMMUNITY—THAT IS CONFERRED BY A PATENT."³⁶

IN THE STRONGEST READING OF THE OPINION, PATENT POLICY IS REDUCED TO SIMPLY ONE MORE FACTOR FOR ANTITRUST ANALYSIS, WHICH IS EFFECTIVELY MADE THE UBER-POLICY...YOU MAY OWN A PATENT, IT SAYS, BUT NONETHELESS, THE COMPETITIONS LAW WILL ALWAYS BE WATCHING.

plication that the antitrust regime sits in a position of supreme oversight over the patent laws.³⁵ Breyer's approach makes clear that, potentially, any anticompetitive consequence thrown off by the patent system could be subject to challenge. To be sure, a patent gives its owner some immunity to antitrust. But "patent and antitrust policies are both relevant in determining the 'scope of the patent monopoly'—and consequently antitrust law immunity—that is conferred by a patent."³⁶

In the strongest reading of the opinion, patent policy is reduced to simply one more factor for antitrust analysis, which is effectively made the uber-policy. That's implied when Breyer says courts should determine liability by "considering traditional antitrust factors such as likely anticompetitive effects, redeeming virtues, market power, and potential offsetting legal considerations present in the circumstances, such as here those related to patents."³⁷ For adherents to a rights model, it must be something to see a property right reduced to merely one of many "potential offsetting legal considerations present in the circumstances."

What we have, ultimately, is a Supreme Court blessing for the use of the competition laws to examine potential excesses of the current patent regime. As such, it represents a powerful victory for a method that uses the antitrust laws to improve the functioning of the patent system, not from within the law itself, but by external measurement of the law's consequences for the economy. You may own a patent, it says, but nonetheless, the competitions law will always be watching. ▲

1. Isidor and Seville Sulzbacher Professor of Law, Columbia Law School. My thanks to Scott Hemphill for comments, and to Bill Toth for timely research assistance.

2. U.S. Dep't of Justice & Fed. Trade Comm'n, Antitrust Guidelines for the Licensing of Intellectual Property § 1.0 (1995).

3. Other examples include Herbert Hovenkamp, Antitrust and Innovation: Where We Are and Where We Should Be Going, 77 Antitrust L.J. 749, 750 (2011); Mark A. Lemley, A New Balance Between IP and Antitrust, 13 Sw. J. L. & Trade Am. 237 (2007); Harry First, Microsoft and the Evolution of the Intellectual Property Concept, 2006 Wisc. L. Rev. 1369 (2006); Michael

- A. Carrier, *Unraveling the Patent-Antitrust Paradox*, 150 U. Pa. L. Rev. 761 (2002); Daniel J. Gifford, *The Antitrust/intellectual Property Interface: An Emerging Solution to an Intractable Problem*, 31 Hofstra L. Rev. 363 (2002).
4. A prominent suggestion for more experimentalism in patent is Lisa Larrimore Ouellette, *Patent Experimentalism* (working paper 2013); see also Tim Wu, *Bad Uses of Good Laws*, N.Y. Times, Oct. 10, 2012.
 5. For a definitive recent work on legal pragmatism, see Richard A. Posner, *LAW PRAGMATISM AND DEMOCRACY*, (2003),
 6. John Dewey, *DEMOCRACY AND EDUCATION: AN INTRODUCTION TO THE PHILOSOPHY OF EDUCATION* 74 (Dover Publ'ns, Inc. 2004) (1916).
 7. *THE MORAL WRITINGS OF JOHN DEWEY* 261 (James Gouinlock, ed.) [need year of the edition]
 8. John Dewey, *THE PUBLIC AND ITS PROBLEMS* 202-03 (Swallow Press Books 1954) (1927).
 9. Cf. William H. Simon, *Toyota Jurisprudence: Legal Theory and Rolling Rule Regimes*, in *LAW AND NEW GOVERNANCE IN THE EU AND THE US*, available at http://www.law.columbia.edu/faculty/fac_resources/faculty_lunch?exclusive=filemgr.download&file_id=9477&rtcontentdisposition=filename%3DSimon.pdf.
 10. See, e.g., Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267, (1998); William Simon & Charles Sabel, *Minimalism and Experimentalism in the Administrative State*, 100 Georgetown L. J. 53 (2010); James S. Liebman & Charles F. Sabel, *A Public Laboratory Dewey Barely Imagined: The Emerging Model of School Governance and Legal Reform*, 28 N.Y.U. REV. L. & SOC. CHANGE 183 (2003); Michael C. Dorf, Charles F. Sabel, *Drug Treatment Courts and Emergent Experimentalist Government*, 53 VAND. L. REV. 831 (2000)
 11. William Simon & Charles Sabel, *Minimalism and Experimentalism in the Welfare State*, 100 Georgetown L. J. 53 (2010).
 12. As Posner writes, pragmatism is “the secret story of our courts as of our political system in general.” *Id.* at 333.
 13. Thurman W. Arnold, *THE BOTTLENECKS OF BUSINESS* 3 (1940).
 14. *FTC. V. Actavis, Inc.*, 133 S. Ct. 2223, 2238 (2013).
 15. A older, alternative position might say that the point of antitrust is to
 16. *RESEARCH HANDBOOK ON INTERNATIONAL COMPETITION LAW* 92 (ed. Ariel Ezrachi) (2012).
 17. *United States v. Aluminum Co. of Am.*, 148 F.2d 416, 432 (2d Cir. 1945) (“To read the passage as demanding any ‘specific,’ intent, makes nonsense of it, for no monopolist monopolizes unconscious of what he is doing.”).
 18. *THE NEXT ANTITRUST AGENDA: THE AMERICAN ANTITRUST INSTITUTE’S TRANSITION REPORT ON COMPETITION POLICY* 58 (Albert Foer ed., 2008) (citing U.S. DEP’T OF JUSTICE, ANTITRUST DIV., WORKLOAD STATISTICS for the proposition that the DOJ has brought no monopolization cases under the Bush Administration) available at http://www.antitrustinstitute.org/files/Monopoly%20Chapter%20from%20%20AAI%20Transition%20Report_100520082111.pdf.
 19. See generally Nicholas Levy, *EU Merger Control: A Brief History*, CLEARY GOTTLEIB STEIN & HAMILTON (2004) http://www.cgsh.com/files/Publication/39346756-bc80-4fd2-9584-f358ffc72239/Presentation/PublicationAttachment/05b61f33-f646-4c9e-a7ec-f6b8b0bfce4f/CGSH_CGSH_Paper_IBC_Conference_EU_Merger_Control_-_A_Brief_History.pdf. [ultimately, I think we’ll want to dive into the sources cited therein, rather than to this document itself]
 20. See *id.* at 15-16 and citations therein.
 21. Council Regulation 139/2004 of 20 January 2004 on the Control of Concentrations between Undertakings, 2004 O.J. (L 24). For further discussion of the 2004 changes to the European merger control regulations, see Werner Berg, *New EC Merger Regulation: A First Assessment of its Practical Impact*, *The Symposium on European Competition Law*, 24 Nw. J. Int’l L. & Bus. 683 (2003-2004) available at <http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1586&context=njillb>.
 22. E.g. Malcolm B. Coate. *The Use of Natural Experiments in Merger Analysis*, Federal Trade Commission (July 19, 2012).
 23. For a recent suggestion of how this might be accomplished, see Lisa Larrimore Ouellette, *Patent Experimentalism* (working paper 2013), *supra*.
 24. Cf. Robert Merges, *Justifying Intellectual Property* (2010).
 25. A full history can be found in Tim Wu, *Copyright’s Communications Policy*, U. Mich. L. Rev., and Jessica Litman

26. See Robert D. Swanson, *Implementing the E.U. Unified Patent Court: Lessons from the Federal Circuit*, 9 B.Y.U. INT'L L. & MGMT. REV. 169 (2013) and citations therein, particularly to Council Draft Agreement 16741/11, Draft Agreement on a Unified Patent Court, of 11 November 2011, Annex.
27. European Patent Convention, Oct. 5, 1973, 1065 U.N.T.S. 199.
28. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 104 S. Ct. 774, 78 L. Ed. 2d 574 (1984).
29. Cf. Charles River Bridge v. Warren Bridge, 36 US 420 (1837).
30. The last was Illinois Tool Works Inc. v. Indep. Ink, Inc., 547 U.S. 28 (2006) (holding that the existence of a patent on a product tied to an unpatented product is not sufficient to establish the market power necessary to find illegal tying).
31. Eldred v. Ashcroft, 537 U.S. 186, 242 (2003) (Breyer J., dissenting).
32. Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281 (1970)
33. The closest he came was in Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 548 U.S. 124, 138 (2006) (“a decision from this generalist Court could contribute to the important ongoing debate, among both specialists and generalists, as to whether the patent system, as currently administered and enforced, adequately reflects the “careful balance” that “the federal patent laws ... embod[y].”) (Breyer, J., dissenting).
34. F.T.C. v. Actavis, Inc., 133 S. Ct. 2223, 2237.
35. *Id.* at 2227 (2013).
36. *Id.* at 2231.
37. *Id.*

Whither Symmetry? Antitrust Analysis of Intellectual Property Rights at the FTC and DOJ

BY JOSHUA D. WRIGHT & DOUGLAS H. GINSBURG ¹

In modern antitrust law, intellectual property rights (IPRs) are treated like all other forms of property. Beginning with the Department of Justice Antitrust Division's repudiation of the "Nine No-No's" for patent licensing more than thirty years ago, the US antitrust enforcement agencies and the courts have adhered to a largely symmetrical approach to antitrust enforcement involving all kinds of property. The principle of symmetry was made express in the 1995 Guidelines for the licensing of intellectual property, issued jointly by the Department of Justice and the Federal Trade Commission.

In this article we consider recent instances in which the antitrust agencies have departed from the principle of symmetry in enforcement actions and in policy statements to the effect that injunctive relief is inappropriate in some IPR settings, that IPR-related mergers require special review, and that breach of an IPR-enmeshed contract gives rise to antitrust liability. All reduce to the same unsupported proposition: IPRs are inherently different from other property rights and, for antitrust purposes, inherently suspect. We call upon the FTC and the DOJ to reaffirm the symmetry principle of the 1995 Guidelines, which provides a consistent and predictable theoretical framework for antitrust challenges involving intellectual property, just as antitrust functions for all other sorts of property. Consumers need a special antitrust doctrine for intellectual property no more than they need antitrust rules specific to grocers.

The modern approach of antitrust law to intellectual property rights (IPRs) is in parity with the contemporary approach of antitrust law to all other forms of property.² This has not always been the case, however. The "inhospitality tradition" of the Federal Trade Commission and of the Antitrust Division of the Department of Justice, as applied to IPRs, culminated decades ago in the "Nine No-No's," a set of formalistic, *per se* prohibitions of various patent arrangements. Agencies and courts have since then adopted a largely symmetrical approach to antitrust enforcement involving IPRs and other forms of property.

Several other doctrinal developments support the principle of symmetry in antitrust enforcement concerning intellectual and other property rights. Key among these is rejection of presumption that IPRs are "inherently suspect." That view, now discredited, arises from a misperceived contradiction between antitrust law and IPRs: antitrust constrains monopoly power, whereas IPRs confer monopoly power. This false opposition was repudiated by the Supreme Court in *Illinois Tool Works*, and with good reason;³ that IPRs necessarily confer monopoly power is inconsistent with sound economic theory and demonstrably false. The proposition that antitrust enforcement and IPRs conflict has been overtaken by the realization that each regime spurs dynamic competition.

The FTC/DOJ 1995 Guidelines on licensing IPR provide that the "[a]gencies apply the same general antitrust principles to conduct involving intellectual property that they apply to conduct involving any other form of tangible or intangible property,"⁴ but they also recognize that such symmetry does not mean they should ignore the unique features of IPRs. Rather, the agencies have sensibly incorporated economic theory and em-

pirical data regarding these unique characteristics into the standard analysis. No departure from the principle of symmetry is, or has been, necessary to do that.⁵

We consider in this article whether the antitrust agencies are at times departing from the modern conventions regarding IPRs outlined above. Antitrust academia does not lack in advocates for such a departure. A good deal of recent antitrust scholarship calls for more interventionist antitrust policy regarding IPRs – sometimes even expressly challenging the symmetry principle and calling instead for IP-specific antitrust treatment. Authors

WE CONSIDER IN THIS ARTICLE WHETHER THE ANTITRUST AGENCIES ARE AT TIMES DEPARTING FROM THE MODERN CONVENTIONS REGARDING IPRs OUTLINED ABOVE. ANTITRUST ACADEMIA DOES NOT LACK IN ADVOCATES FOR SUCH A DEPARTURE.... WE FIND TROUBLING A NUMBER OF RECENT ENFORCEMENT ACTIONS THAT DEPART FROM THE PRINCIPLE OF SYMMETRY IN ANALYZING ANTITRUST PROBLEMS INVOLVING IPRs.

expressing this view generally claim the “probabilistic” nature of patent rights induces more opportunistic or anticompetitive behavior than do rights in other types of property.⁶ Some claim the optimal level of antitrust enforcement depends upon the strength of the patent rights associated with each particular industry. For example, Professor Mark Lemley argues the antitrust laws should be strong where IPRs are also strong.⁷ In this view, antitrust acts as an *ad hoc* counterweight to IP; the need for antitrust enforcement “depends on the industry in question and the nature of the invention.”⁸

We find troubling a number of recent enforcement actions that depart from the principle of symmetry in analyzing antitrust problems involving IPRs. In particular, we examine in this article recent agency testimony and enforcement actions discouraging holders of standard essential patents (SEPs) from pursuing preliminary injunctions or exclusion orders, enforcement actions arguing that breach of a FRAND commitment constitutes an antitrust violation, and merger review involving SEPs.

These agency actions may evince a pattern of increasing hostility to IPRs. They certainly represent a departure from the modern approach and imply changes in antitrust doctrine and enforcement more generally. Two such changes bear special emphasis here. The first is a rejection of the symmetry principle in favor of expanding the application of antitrust to IPRs in way that does not apply to other types of property. The second is an extension of antitrust liability that blurs the fundamental distinction between ordinary contract law and antitrust law.

I. STANDARD ANTITRUST ANALYSIS INVOLVING IPRs

The symmetry principle holds antitrust limits the operation of IPRs in the same way it limits other rights in property. Symmetry cabins IPR-related antitrust claims and defenses alike. It implies an antitrust claim based upon the use of one’s IPR is no more suspect than a claim arising from the use of any other form of property; it also implies an antitrust claim cannot be defended on the ground that the use of an IPR is inherently less suspect than the use of some other form of property. The US antitrust agencies have recognized the “symmetry limitation” to antitrust claims in the IP Licensing Guidelines,⁹ and the D.C. Circuit endorsed the corollary limitation to defenses in *United States v. Microsoft*, wherein it observed Microsoft’s claim to an “unfettered right to use its intellectual property as it wishes” was “no more correct than the proposition that use of one’s personal property, such as a baseball bat, cannot give rise to tort liability.”¹⁰

An obvious corollary of the symmetry principle is that antitrust claims involving IPRs require no special antitrust presumptions. This does not mean conduct involving IPRs cannot run afoul of antitrust presumptions regarding inherently suspect or presumptively unlawful activity. For example, naked price-fixing is correctly held unlawful *per se* whether the underlying assets involve intellectual or other types of property rights.¹¹ Such a presumption, however, is based upon the anticompetitive effect, not the form, of the conduct. That the conduct involves the transfer, acquisition, or licensing of a particular kind of property, such as an IPR, is irrelevant to the presumption.¹² At the same time, symmetry in the antitrust analysis applied to intellectual property and to other property does not imply or require one to ignore the specific factual or institutional features of the underlying property rights.¹³

We now illustrate operation of the symmetry principle across a representative sample of business conduct in order to highlight recent departures from that principle. Consider licensing agreements, which can be challenged as either collusive or exclusionary under the antitrust laws. Consistent with the symmetry principle, the standard approach in examining an allegedly collusive patent agreement is to focus upon whether the patents are substitutes or complements. Thus, in *Summit Technology, Inc. & VISX, Inc.*,¹⁴ the FTC, applying standard antitrust principles, acknowledged that horizontal relationships involving IPRs can generate efficiencies — but also warned that agreements facilitating naked price fixing or market division would violate the antitrust laws.¹⁵ Summit and VISX had agreed to pool their patents for their rival laser eye surgery technologies, require licensees to pay \$250 to the pool each time a procedure was performed using either firm’s technology, and then split the proceeds according to a specified formula.¹⁶ The FTC successfully argued the fee established a *de facto* price floor, meaning the companies had colluded to charge a minimum fee of \$250 per procedure.¹⁷

In contrast, licensing agreements that include complementary IPR are unlikely to raise antitrust concerns. Consider the DOJ Business Review Letters concerning the various patent pools covering DVD-video, DVD-ROM, MPEG-2, and 3G standards.¹⁸ These pools included only patents “essential” to the relevant standard, which usually meant complementary patents. This focus on complementarity is not unique to IPRs; rather, it is the symmetrical application to IPR agreements of standard antitrust analysis of horizontal relationships, such as joint ventures.¹⁹

Vertical IP licensing agreements are also subject to the same analytical approach applied to vertical agreements involving other property rights. In *Illinois Tool Works, Inc. v. Independent Ink, Inc.*,²⁰ the Supreme Court, undertaking a “reexamination of the presumption of *per se* illegality of a tying arrangement involving a patented product,”²¹ rejected the “patent-equals-market-power presumption” upon which it was based, and thereby joined the longstanding consensus among antitrust scholars,²² the Congress,²³ and the antitrust agencies²⁴ that patents do not ordinarily confer monopoly power. *Independent Ink* is an avowed endorsement of symmetrical antitrust analysis for vertical IPR licensing agreements.

In fact, the symmetry principle holds across the panoply of IPR licensing arrangements. The Guidelines note that although exclusive dealing agreements involving IPR are often pro-competitive, they “may anticompetitively foreclose access to, or increase competitors’ costs of obtaining, important inputs, or facilitate coordination to raise price or reduce output.”²⁵ This is precisely the anticompetitive concern with exclusive dealing and tying contracts involving other forms of property.²⁶ The relevant factors are identical for all types of property: the percentage of inputs potentially foreclosed from rivals by the licensing arrangement or exclusive contract, the duration of the license, the ability of rivals to realign supply contracts, and barriers to entry.²⁷

Refusals to license IPRs also follow the symmetry principle. The antitrust laws have long recognized that a firm generally may, in its “own independent discretion, decline to deal with another firm.”²⁸ Courts have been reluctant to use the antitrust laws to obligate even a firm with monopoly power to deal with a rival, and in *Trinko*,²⁹ the Supreme Court placed the refusal-to-deal liability in *Aspen Skiing*³⁰ “at or near the outer boundary of §2 liability.”³¹ Patent law takes a similar approach.³² Lower courts have faithfully applied the symmetry principle, treating refusals to deal as virtually *per se* lawful³³ or granting the IP holder a presumption of legitimacy.³⁴

Finally, the symmetry principle applies to horizontal merger analysis involving IPRs. The presence of IPRs does not alter the key question identified by the agencies in the 2010 Horizontal Merger Guidelines, whether the merger creates market power that would not otherwise exist. This question requires analysis of the pre- and post-merger incentives and abilities of the merging firms; IPRs may be relevant to the analysis but only to the extent that any other property would be. Indeed, the only reference to IPRs in the Horizontal Merger Guidelines emphasizes that their combination may affect the merged firm’s ability to appropriate the gains from, and thereby its incentive to engage in, innovative activities.³⁵

...MODERN ANTITRUST ANALYSIS
AT EVERY TURN REFLECTS THE
SYMMETRY PRINCIPLE ENDORSED
IN THE 1995 ANTITRUST IP
GUIDELINES....RECENT ACTIONS BY
THE ANTITRUST ENFORCEMENT
AGENCIES – INCLUDING ENFORCEMENT
ACTIONS, TESTIMONY, AND SPEECHES
– SUGGEST THE BEGINNING OF
WHAT COULD BE A WHOLESALE
DEPARTURE FROM THE PRINCIPLE
OF SYMMETRY DESCRIBED ABOVE.

In sum, modern antitrust analysis at every turn reflects the symmetry principle endorsed in the 1995 Antitrust IP Guidelines. We next examine whether recent agency actions and statements depart from that principle.

II. IS THE SYMMETRY PRINCIPLE UNDER ATTACK AT THE ANTITRUST AGENCIES?

Recent actions by the antitrust enforcement agencies – including enforcement actions, testimony, and speeches – suggest the beginning of what could be a wholesale departure from the principle of symmetry

described above. More specifically, they evince a drift toward *ad hoc* antitrust treatment of IPRs in myriad settings, including the pursuit of preliminary injunctions or exclusion orders by holders of SEPs; the refusal of SEP holders to license them; merger analysis of transactions involving SEPs or the aggregation of patent rights by patent assertion entities (PAEs); and breaches of contract, especially FRAND contracts with standard setting organizations (SSOs).

Consider first the hostility of the enforcement agencies to SEP holders seeking injunctive relief. To be sure, there is scholarly work exploring the possibility of patent hold-up – usually perfected by the patentee seeking of an injunction in order to extract “supra-competitive” royalties³⁶ – but this only indicates that an injunction against infringement of a patent *could* be inefficient and potentially anticompetitive. It is a far cry from demonstrating, as a matter of either economic theory or empirical data, that the enforcement of a presumptively valid property right is inherently anticompetitive. The reflexive position that an SEP holder violates the antitrust laws simply by seeking an injunction to vindicate its right clearly departs from the symmetry principle.³⁷ Antitrust law does not prohibit the holder of any other right from seeking an injunction to vindicate that right and, of course, it is hardly uncommon for property rights holders to do so.

In many SSOs, the availability of injunctive relief against an infringer is very likely part of the background understanding between the SSO and its members; in fact, the right to an injunction likely accounted in part for the patent owners' decisions to join the SSO and contribute technologies under a F/RAND commitment. "Hold-up" arguments against a patentee pursuing an injunction often neglect the "reverse holdup" problem: weakening the availability of injunctive relief may undermine the incentive of SSOs to negotiate in good faith with patent holders. A key component of property, including intellectual property, is the right of the owner to exclude others from use of the property, which enables clear assignment of property rights and facilitates economic exchange. Like so many situation-specific departures from established antitrust law, "reforms" to fix the SEP hold-up problem risk reducing welfare by raising an overlooked cost.³⁸

Despite this risk, however, the antitrust agencies are increasingly opposing injunctions for aggrieved SEP holders. Recently, for example, the FTC submitted an *amicus* brief to the Federal Circuit in support of a district court's denial of injunctive relief to a RAND-encumbered holder of an SEP.³⁹ Similarly, in June 2012, the FTC encouraged the International Trade Commission to apply the "public interest standard" of Section 337 of the Tariff Act of 1930 in a manner

that would preclude granting an injunction to an SEP holder on the ground that consumers would be harmed thereby.⁴⁰ The Commission has even encouraged the Congress to take up this narrow remedial issue, abdicating the Commission's role as economic analyst in favor of proposing a simplistic legislative change.⁴¹

The DOJ has followed suit. In its Joint Statement on Remedies for Standards-Essential Patents Subject to Voluntary F/RAND Commitments, the DOJ endorsed the view that an exclusion order should generally not be granted because in a particular case "[a] decision maker could conclude that the holder of a F/RAND encumbered SEP had attempted to use an exclusion order to pressure an implementer of a standard to accept more onerous licensing terms than the patent holder would be entitled to receive consistent with the F/RAND commitment."⁴² The DOJ would allow exclusion orders "where the putative licensee is unable or refuses to take a F/RAND license and is acting outside the scope of the patent holder's commitment to license on F/RAND terms" or when the putative licensee is not subject to the jurisdiction of a court that could award damages.⁴³

Neither economic theory nor empirical evidence supports this novel and discriminatory view of FRAND-encumbered SEP holders as unworthy of an injunction to vindicate their rights. Economic theory recognizes the pro-competitive as well as the anticompetitive uses of injunctions. The FTC/DOJ position opposing injunctions for SEP holders in almost all circumstances relies upon an inexplicable presumption of net anticompetitive harm. That presumption is at odds with the conventional antitrust treatment of the exercise of property rights, and is thus in significant tension with the symmetry principle in the 1995 Antitrust-IP Guidelines.⁴⁴

The FTC's opposition to the issuance of injunctions goes beyond mere advocacy. Consider two recent enforcement actions by the agency alleging that a F/RAND-encumbered SEP holder violated Section 5 of the FTC Act by seeking an injunction.⁴⁵ In *Bosch*, the FTC alleged an SEP holder's pursuit of an injunction was an unfair method of competition.⁴⁶ In *Motorola*, the FTC alleged the company "breached its FRAND obligations

NEITHER ECONOMIC THEORY NOR EMPIRICAL EVIDENCE SUPPORTS THIS NOVEL AND DISCRIMINATORY VIEW OF FRAND-ENCUMBERED SEP HOLDERS AS UNWORTHY OF AN INJUNCTION TO VINDICATE THEIR RIGHTS. ECONOMIC THEORY RECOGNIZES THE PRO-COMPETITIVE AS WELL AS THE ANTICOMPETITIVE USES OF INJUNCTIONS.

WE THINK THIS SPARSE INTELLECTUAL FOUNDATION IS INSUFFICIENT TO SUPPORT THE SHIFT AWAY FROM THE DECADES-LONG CONVENTION RESPECTING THE SYMMETRY BETWEEN IPR AND NON-IPR ANTITRUST ANALYSIS.

by seeking to enjoin and exclude implementers of its SEPs;” that after its acquisition of Motorola, “Google used these threats of exclusion orders and injunctions to enhance its bargaining leverage against willing licensees;” and that “Motorola filed, and Google prosecuted, patent infringement claims before the United States International Trade Commission.”⁴⁷ These complaints

and consent orders, taken together, logically and necessarily depend upon the presumption that protecting a valid SEP against infringement by obtaining injunctive relief is itself anticompetitive.

There is similar evidence suggesting the IPR-specific, asymmetric view extends also to merger review. The consent decree in *Bosch* arose from the FTC’s review of the proposed acquisition of SPX Service Solutions. SPX Service Solutions, an SEP holder, had sued competitors for infringing patents essential to the practice of a standard. This, of course, is insufficient to condemn SPX Service Solutions’ litigation strategy: one cannot know, without more information, whether the injunctions would have proven anticompetitive. There is nothing unusual about a competition agency’s uncovering and prosecuting additional unlawful activity during its investigation of a merger, but in this instance the resulting consent order seems to be an incremental extension of the FTC’s asymmetric stance against patentees with SEPs, if not against IPRs more generally.

Signs of asymmetry are also found at the DOJ. On February 13, 2012, the Antitrust Division issued a closing statement in its investigation of (1) Google Inc.’s acquisition of Motorola Mobility Holdings Inc.; (2) Apple Inc., Microsoft Corp., and Research in Motion Ltd.’s (RIM’s) acquisition of patents from Nortel Networks Corp.; and (3) Apple Inc.’s acquisition of patents from Novell Inc.⁴⁸ The acquisitions all involved patents that would assist the acquiring firm in the development of cellular phone technology, including smartphones and operating systems for those phones. The DOJ had expressed concern that the acquiring companies would use the SEPs they obtained to hold up rivals, to the detriment of competition and innovation.

The DOJ concluded its investigation by declaring that none of the acquisitions was likely to lessen competition substantially, but this nominal approval concealed a radical concession by the firms. In determining it was proper to close the investigations, the DOJ “took into account the fact that during the pendency of these investigations, Apple, Google and Microsoft each made public statements explaining their respective SEP licensing practices.”⁴⁹ Specifically, Apple and Microsoft agreed not to seek injunctions for violations of their licenses they were acquiring. Google’s commitment was “less clear” to the DOJ, for which reason it “continue[d] to have concerns about the potential inappropriate use of SEPs to disrupt competition.”⁵⁰ Concluding its analysis, the DOJ emphasized its obligation to balance the “rightful exercise of patent rights” against “the anticompetitive use of those rights;”⁵¹ noticeably missing was any acknowledgement of the potential pro-competitive benefits to SSOs of injunctions against members that infringe SEPs. The DOJ’s shift in IPR enforcement policy invites the inference that the “voluntary” commitments were in fact a condition laid down by the DOJ for getting the acquisitions approved.

Each of these departures from standard antitrust analysis rests squarely upon the presumption – certainly not the demonstration – that injunctions granted to SEP holders are inherently anticompetitive or likely to be so. This presumption significantly expands the scope of antitrust liability across a variety of dynamic commercial sectors. Yet the most common defense of the presumption illustrates the second serious problem with

the agencies' new policy: it blurs the distinction between contract law and antitrust law.

The most common defense of the presumption is that seeking the injunction itself amounts to a breach of contract,⁵² which is, in turn, the crux of the antitrust violation.⁵³ At least two immediate problems arise. First, mere breach of contract is generally not an antitrust violation: even in patent holdups, federal courts require some additional conduct, such as deception in the standard-setting process, to ground an antitrust violation.⁵⁴ Second, the “injunction-seeking as breach-of-contract” theory depends upon the assumption that a F/RAND commitment comprises an implicit agreement not to seek an injunction. This is far from clear, however. Even a cursory examination of actual industry practice suggests the opposite: no SSO appears expressly to disallow injunctions,⁵⁵ and some SSOs appear expressly to have considered and rejected such a rule. Prohibiting the use of an injunction, as noted before, would subject the SEP holder to reverse hold up in which the potential licensee refuses to pay a royalty that in fact is fair, reasonable and non-discriminatory.

Given the agencies' newfound hostility to IPRs, it should come as no surprise that the FTC has also dubbed some garden-variety breaches of contract as antitrust violations. A glaring example is *N-Data*, in which the Commission alleged that departure from a contractual commitment to an SSO violated Section 5 of the FTC Act without proof that deceptive conduct accounted for the SSO adopting its technology.⁵⁶ Over dissenting statements from Chairman Majoras and Commissioner Kovacic,⁵⁷ the FTC ruled that departure from a contractual commitment to an SSO, standing alone, was enough to violate Section 5. At least one DOJ official has supported exploration of the theory that similar conduct violates Section 2 of the Sherman Act.⁵⁸

One defense of the proposition that antitrust is an appropriate regime for governing contractual disputes involving SSOs is that an SSO is best conceived of as a collaboration among competitors that have entered into a *de facto* agreement with antitrust authorities by which the authorities allow collusive interaction (in the form of standardization) in exchange for tougher antitrust scrutiny. But there is no evidence of such a *de facto* agreement, much less an agreement to depart from standard antitrust principles. Further, there is no empirical evidence supporting the proposition that breach of an SSO contract – even one resulting in higher royalty rates – has economic effects similar to the collusive interaction between rivals condemned by the antitrust laws. And, as we have noted, courts have uniformly rejected this view when interpreting and applying the Sherman Act; SSOs have been routinely considered pro-competitive.

FOR THE ANTITRUST AGENCIES TO FALL PREY TO SUCH TEMPTATION, HOWEVER, TEMPTS AN OBSERVER TO DESPAIR AT THE ANTITRUST PROJECT; IT SOMETIMES APPEARS ANTITRUST ENFORCERS ARE ESPECIALLY PRONE TO REPEATING HISTORICAL MISTAKES. YET THIS MISTAKE IS NOT ONLY NOT ANCIENT, ITS REPUDIATION IS RECENT AND SHOULD BE FRESH IN MEMORY.

In any event, the shifting approach to IPRs should be clear from the FTC and DOJ testimony in favor of restricting the availability of injunctions, from the FTC's enforcement actions attacking the pursuit of an injunction as a violation of the antitrust laws, and from the extension of Section 5 to condemn breach of a F/RAND contract with an SSO. The analytical basis for this shift rests upon two points. The first is the presumption that injunctions against infringement of a patent are inherently anticompetitive, particularly with regard to F/RAND-encumbered SEPs. The second is the inference that a FRAND commitment entails a promise not to pursue injunctive relief for infringement. We think this sparse intellectual foundation is insufficient to

support the shift away from the decades-long convention respecting the symmetry between IPR and non-IPR antitrust analysis.

III. CONCLUSIONS

Each of the arguments against symmetry for IPRs is ultimately unpersuasive. Recent agency insistence that injunctive relief is inappropriate in some IPR settings, or that IPR-related mergers require special review, or that IPR-enmeshed contracts give rise to antitrust liability for ordinary breaches of contract, all reduce to the same unsupported proposition: IPRs are inherently different from other property rights and, for antitrust purposes, inherently suspect.

We disagree. The notion that IPRs present a distinct theoretical problem for antitrust is hardly unprecedented; industries from railroads to fashion to baseball have alleged justifications for unprincipled exceptions to generally applicable antitrust rules. We can understand, if not condone, those attempts by private parties; a litigant is famously entitled to its self-serving positions, and the courts and agencies alike have nearly uniformly rejected these requests for exceptions. For the antitrust agencies to fall prey to such temptation, however, tempts an observer to despair at the antitrust project; it sometimes appears antitrust enforcers are especially prone to repeating historical mistakes. Yet this mistake is not only not ancient, its repudiation is recent and should be fresh in memory. We hope the FTC and DOJ will reaffirm the symmetry principle in the 1995 Guidelines, note that the symmetry principle provides a consistent and predictable theoretical framework for antitrust challenges involving intellectual property, just as antitrust functions for all other sorts of property. Consumers need an antitrust doctrine for intellectual property no more than they needed antitrust rules specific to grocers. ▲

1. Joshua D. Wright is Commissioner, Federal Trade Commission, and Professor of Law (on leave) at George Mason University School of Law. The views expressed here are the author's and do not necessarily reflect those of the Commission or any other Commissioner. Douglas H. Ginsburg is Senior Circuit Judge, United States Court of Appeals for the District of Columbia Circuit, and Professor of Law at George Mason University. The authors thank Keith Hylton, Bruce Kobayashi, and Timothy J. Muris for valuable comments on an earlier draft and Samantha Morelli, Kristin Stortini, and Zander Li for helpful research assistance.

2. We exclude from the scope of our analysis in this paper antitrust scrutiny of conduct preceding the issuance of IPRs.

3. See *Illinois Tool Works Inc. v. Indep. Ink, Inc.*, 547 US 28 (2006), *rev'g* 396 F.3d 1342 (Fed. Cir. 2005).

4. US DEP'T OF JUSTICE & FED. TRADE COMM'N, ANTITRUST GUIDELINES FOR THE LICENSING OF INTELLECTUAL PROPERTY § 2.1 (1995), *reprinted in* 4 Trade Reg. Rep. (CCH) ¶ 13,132, *available at* <http://www.usdoj.gov/atr/public/guidelines/0558.pdf> [hereinafter ANTITRUST-IP GUIDELINES].

5. *Id.* (“That is not to say that intellectual property is in all respects the same as any other form of property. Intellectual property has important characteristics, such as ease of misappropriation, that distinguish it from many other forms of property. These characteristics can be taken into account by standard antitrust analysis, however, and do not require the application of fundamentally different principles.”).

6. The term “probabilistic” reflects the generally accepted proposition that the validity of patent is, on average, less certain than the validity of the title to a piece of real or personal property. See, e.g., Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSP. 75 (2005).

7. Mark A. Lemley, *New Balance Between IP and Antitrust*, 13 SW. J.L. & TRADE AM. 237, 237 (2007). The argument is grounded in the tradeoff between the benefits, on the one hand, of innovation and dynamic competition and, on the other hand,

the deadweight loss associated with monopoly. One difficulty with calibrating antitrust policy on an industry-by-industry basis based upon a presumed positive or “inverted U-shape” relationship between static product market competition and innovation is that neither economic theory nor the empirical record supports such a presumption. See Douglas H. Ginsburg & Joshua D. Wright, *Dynamic Analysis and the Limits of Antitrust Institutions*, 78 ANTITRUST L.J. 1, 4-5 (2012).

8. Lemley, *supra* note 6, at 237.

9. ANTITRUST-IP GUIDELINES, *supra* note 3, § 2.1, at 3 (discussing that the special characteristics of IP “can be taken into account by standard antitrust analysis . . . and do not require the application of fundamentally different principles”).

10. United States v. Microsoft Corp., 253 F.3d 34, 63 (2001) (per curiam).

11. See United States v. New Wrinkle, 342 US 371 (1952); ANTITRUST-IP GUIDELINES, *supra* note 3, § 3.4 (explaining the agencies will challenge horizontal agreements involving intellectual property rights as per se violations of the antitrust laws when the “type of restraint is one that has been accorded per se treatment” in other settings). See also *id.* at ex. 7 (describing an agency challenge under the per se rule of “a sham intended to cloak [the] true nature” of a particular licensing agreement).

12. The symmetry principle also applies both to *Walker Process* claims, which require proof not only of the defendant’s fraud upon the patent office but also of all the elements of a Section 2 monopolization claim, see *Walker Process Equip., Inc. v. Food Mach. and Chem. Corp.*, 382 US 172, 179 (1965), and to actions for sham litigation, see *Prof’l Real Estate Investors, Inc. v. Columbia Pictures Indus., Inc.*, 508 US 49, 61 (1993). There is no meaningful difference between the antitrust standard applied to these claims and that faced by a plaintiff alleging monopolization or attempted monopolization by the defendant having fraudulently procured a real property right or other government-granted privilege.

13. See ANTITRUST-IP GUIDELINES, *supra* note 3, § 2.1 (“Intellectual property has important characteristics, such as ease of misappropriation, that distinguish it from many other forms of property. These characteristics can be taken into account by standard antitrust analysis, however, and do not require the application of fundamentally different principles).

14. *In re Summit Tech., Inc.*, 127 F.T.C. 208 (1999).

15. Analysis of Proposed Consent Order to Aid Public Comment, *Summit Tech., Inc.*, FTC File No. 9286, available at <http://www.ftc.gov/os/1998/08/d09286ana.htm>.

16. Complaint at ¶ 12, *In re Summit Tech, Inc.*, 127 F.T.C. 208 (1999), available at <http://www.ftc.gov/os/1998/03/summit.cmp.htm>.

17. *Id.*

18. Letter from Charles A. James, Assistant Attorney Gen., Dep’t of Justice Antitrust Div., to Ky P. Ewing, Esq., Vinson & Elkins L.L.P. (Nov. 12, 2002), available at <http://www.usdoj.gov/atr/public/busreview/200455.pdf>; Letter from Joel I. Klein, Assistant Attorney Gen., Dep’t of Justice Antitrust Div., to Carey R. Ramos, Esq., Paul, Weiss, Rifkind, Wharton & Garrison (June 10, 1999), available at <http://www.usdoj.gov/atr/public/busreview/2485.pdf>; Letter from Joel I. Klein, Assistant Attorney Gen., Dep’t of Justice Antitrust Div., to Garrard R. Beeney, Esq., Sullivan & Cromwell (Dec. 16, 1998), available at <http://www.justice.gov/atr/public/busreview/2121.pdf>; Letter from Joel I. Klein, Assistant Attorney Gen., Dep’t of Justice Antitrust Div., to Garrard R. Beeney, Esq., Sullivan & Cromwell (June 26, 1997), available at <http://www.usdoj.gov/atr/public/busreview/215742.pdf>.

19. FED. TRADE COMM’N & U.S. DEP’T OF JUSTICE, ANTITRUST GUIDELINES FOR COLLABORATION AMONG COMPETITORS § 3.2 (2000), available at <http://www.usdoj.gov/atr/public/busreview/200455.pdf> (“Participants in an efficiency-enhancing integration typically combine, by contract or otherwise, significant capital, technology, or other complementary assets to achieve procompetitive benefits that the participants could not achieve separately.”).

20. *Illinois Tool Works Inc. v. Indep. Ink, Inc.*, 547 US 28 (2006), *rev’g* 396 F.3d 1342 (Fed. Cir. 2005). See generally Joshua D. Wright, *Missed Opportunities in Independent Ink*, 5 CATO SUP. CT. REV. 333 (2006); Bruce H. Kobayashi, *Spilled Ink or Economics Progress: The Supreme Court’s Decision in Illinois Tool Works v. Independent Ink*, 53 ANTITRUST BULL. 5, 12-13 (2008).

21. *Id.* at 38. As others have noted, the Court’s analysis appears to conflate the presumption of market power and a special per se rule of illegality applied to patent ties. See Kobayashi, *supra* note 15, at . The Court is clear, however, that neither the market

power presumption nor the special per se rule applied to patent ties survives the Court's analysis.

22. See, e.g., 10 Philip E. Areeda et al., *Antitrust Law* ¶ 1737a (2d ed. 2004); 1 Herbert Hovenkamp et al., *IP and Antitrust: An Analysis of Antitrust Principles Applied to Intellectual Property Law* § 4.2 (“[A]n intellectual property right does not confer a monopoly.”); Richard A. Posner, *Antitrust Law* 97–98 (2d ed. 2001) (“[M]ost patents confer too little monopoly power to be a proper object of antitrust concern.”).

23. See 35 U.S.C. § 271(d)(4), (5) (requiring proof of market power in the tying product in a patent misuse case), added in 1988.

24. ANTITRUST-IP GUIDELINES, *supra* note 3, § 2.2, at 4.

25. *Id.* § 4.1.2, at 19.

26. See, e.g., Alden F. Abbott & Joshua D. Wright, *Antitrust Analysis of Tying Arrangements and Exclusive Dealing*, in *ANTI-TRUST LAW AND ECONOMICS* 183 (Keith N. Hylton ed., 2010).

27. See, e.g., *United States v. Microsoft Corp.*, 253 F.3d 34, 69-70 (2001).

28. *United States v. Colgate & Co.*, 250 US 300, 307 (1919) (“In the absence of any purpose to create or maintain a monopoly, the [Sherman] act does not restrict the long recognized right of trader or manufacturer engaged in an entirely private business, freely to exercise his own independent discretion as to parties with whom he will deal.”).

29. *Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko*, 540 US 398 (2004).

30. *Aspen Skiing Co. v. Aspen Highlands Skiing Corp.*, 472 US 585 (1985).

31. The Court distinguished *Aspen Skiing* on the grounds that, among other things, (1) the defendant in *Trinko* was not alleged to have ceased a course of dealing with its rivals, (2) the defendant had not refused to sell the relevant services at the retail price it would have charged but for government regulation, and (3) the services at issue were available separately only as a result of government compulsion.

32. See, e.g., *Hartford-Empire Co. v. United States*, 323 US 386, 432-33 (patent owner “has no obligation either to use it or to grant its use to others”), *clarified by* 324 US 570 (1945); *Cont'l Paper Bag Co. v. E. Paper Bag Co.*, 210 US 405, 429 (1908) (“[E]xclusion may be said to have been of the very essence of the right conferred by the patent, as it is the privilege of any owner of property to use or not use it, without question of motive.”); see also *Townshend v. Rockwell Int'l Corp.*, No. C99-0400, 2000 US Dist. LEXIS 5070, at *23 (N.D. Cal. Mar. 28, 2000) (“[A] patent holder is permitted under the antitrust laws to completely exclude others from practicing his or her technology.”).

33. *In re Indep. Serv. Orgs. Antitrust Litig.*, 203 F.3d 1322, 1327-28 (Fed. Cir. 2000) (“[A]bsent exceptional circumstances, a patent may confer the right to exclude competition altogether.”).

34. *Image Technical Services, Inc. v. Eastman Kodak Co.*, 125 F.3d 1195, 1219-20 (9th Cir. 1997). For an economic analysis reconciling the Ninth and Federal Circuit decisions by taking into account the scope of the alleged refusal to deal in each case, see Michelle M. Burtis and Bruce H. Kobayashi, *Why an Original Can Be Better Than a Copy: Intellectual Property, the Antitrust Refusal to Deal, and ISO Antitrust Litigation*, 9 SUP. CT. ECON. REV. 143 (2001).

35. See U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, *HORIZONTAL MERGER GUIDELINES* §10, at 31 (2010), *available at* <http://www.ftc.gov/os/2010/08/100819hmg.pdf> (“The Agencies also consider the ability of the merged firm to appropriate a greater fraction of the benefits resulting from its innovations. Licensing and intellectual property conditions may be important to this enquiry, as they affect the ability of a firm to appropriate the benefits of its innovation.”).

36. See, e.g., Joseph Farrell et al., *Standard Setting, Patents, and Hold-Up*, 74 ANTITRUST L.J. 603, 611-15 (2007). The economic forces at work in such cases of *ex post* opportunism are not, however, patent-specific. Indeed, they were first identified and explored with respect to other types of property. See, e.g., Benjamin Klein, Robert Crawford & Armen A. Alchian, *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, 21 J.L. & ECON. 297 (1978).

37. To be clear, we have no quarrel with the holding in *eBay Inc. v. MercExchange, L.L.C.* 547 US 388, 391 (2006), that whether a permanent injunction should issue in favor of a patent holder is to be determined “consistent with traditional principles of equity,” which require that the plaintiff “demonstrate: (1) that it has suffered an irreparable injury; (2) that remedies available at

law, such as monetary damages, are inadequate to compensate for that injury; (3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) that the public interest would not be disserved by a permanent injunction.” Indeed, the principle of symmetry requires no less. Similarly, the courts properly consider whether a patentee, like any other applicant for a preliminary injunction, has shown that it will probably succeed on the merits of its case. *See Amazon.com, Inc. v. Barnesandnoble.com*, 239 F.3d 1343, 1951 (Fed. Cir. 2001). Rather, our objection is to the idea that a patentee violates the antitrust laws merely by seeking an injunction.

38. *See generally* Joshua D. Wright, Comm’r, Fed. Trade Comm’n, SSOs, FRAND, and Antitrust: Lessons from the Economics of Incomplete Contracts (Sept. 12, 2013), <http://ftc.gov/speeches/wright/130912cpip.pdf>.

39. Brief for Fed. Trade Comm’n as Amicus Curiae Supporting Neither Party at 16, *Apple Inc. v. Motorola, Inc.*, Nos. 2012-1548 & 2012-1549 (Fed. Cir. Dec. 5, 2012) (stating that “[w]hen a patentee makes a FRAND commitment to an SSO, the irreparable harm analysis, balance of harms, and the public interest will, as here, generally militate against an injunction”). Commissioner Wright was not a member of the Commission when the brief was filed; Commissioner Ohlhausen did not vote in favor of its submission.

40. *See* Third Party United States Fed. Trade Comm’n’s Statement on the Public Interest, *In re* Certain Wireless Communication Devices, Portable Music & Data Processing Devices, Computers and Components Thereof, Inv. No. 337-TA-745 (June 6, 2012), www.ftc.gov/os/2012/06/1206ftcwirelesscom.pdf; Third Party United States Fed. Trade Comm’n’s Statement on the Public Interest, *In re* Certain Gaming and Entertainment Consoles, Related Software, and Components Thereof, Inv. No. 337-TA-752 (June 6, 2012), <http://www.ftc.gov/os/2012/06/1206ftcgamingconsole.pdf>. Commissioner Wright was not a member of the Commission when the statement was filed.

41. Fed. Trade Comm’n, Prepared Statement of the Fed. Trade Comm’n before S. Comm. on the Judiciary, Concerning Oversight on the Competition of Exclusion Orders to Enforce Standard-Essential Patents at 2 (July 11, 2012), *available at* <http://www.ftc.gov/os/testimony/120711standardpatents.pdf>

42. UNITED STATES DEP’T OF JUSTICE & UNITED STATES PATENT & TRADEMARK OFFICE, POLICY STATEMENT ON REMEDIES FOR STANDARDS-ESSENTIAL PATENTS SUBJECT TO VOLUNTARY F/RAND COMMITMENTS, at 6 (2013), *available at* http://www.uspto.gov/about/offices/ogc/Final_DOJ-PTO_Policy_Statement_on_FRAND_SEPs_1-8-13.pdf.

43. *Id.*

44. We also aware of recent instances in which the FTC and DOJ continue to endorse the symmetry principle, at least rhetorically. *See, e.g.*, Renata Hesse, Deputy Assistant Attorney Gen., United States Dep’t of Justice, IP, Antitrust, and Looking Back on the Last Four Years (Feb. 8, 2013) (“[T]he Antitrust Division applies the same general antitrust principles to mergers and conduct matters involving intellectual property that it applies to any other type of property. That stance will continue.”).

45. Commissioner Wright did not participate in either of these decisions.

46. Complaint at ¶ 20, *In re* Robert Bosch GmbH, Docket No. C-4377, File No. 121-0081, 2012 WL 5944820 (Fed. Trade Comm’n, Nov. 21, 2012).

47. Complaint at ¶¶ 25-26, *In re* Motorola Mobility LLC, Docket No. C-4410, File No. 121-0120, 2013 WL 3944149 (Fed. Trade Comm’n, July 23, 2013).

48. Press Release, United States Dep’t of Justice, Statement of the Department of Justice’s Antitrust Division on Its Decision to Close Its Investigations of Google Inc.’s Acquisition of Motorola Mobility Holdings Inc. and the Acquisitions of Certain Patents by Apple Inc., Microsoft Corp. and Research in Motion Ltd. (Feb. 13, 2012), *available at* http://www.justice.gov/atr/public/press_releases/2012/280190.pdf.

49. *Id.*

50. *Id.*

51. *Id.*

52. *Apple, Inc. v. Motorola, Inc.*, 869 F. Supp. 2d 901, 914 (N.D. Ill. 2012) (“By committing to license its patents on FRAND terms, [the patent holder] committed to license the [patent-in-suit] to anyone willing to pay a FRAND royalty and thus implicitly

acknowledged that a royalty is adequate compensation for a license to use that patent.”).

53. This theory has been asserted primarily as a violation of Section 5 of the FTC Act. *See, e.g., In re Motorola Mobility LLC*, Docket No. C-4410, File No. 121-0120, 2013 WL 3944149 (Fed. Trade Comm’n, July 23, 2013); *In re Negotiated Data Solutions, LLC*, FTC File No. 051-0094 (Jan. 23, 2008), *available at* <http://www.ftc.gov/os/caselist/0510094/080122statement.pdf>. *Cf. Hesse, supra* note 43, at 21 (indicating the DOJ’s interest in examining whether Section 2 of the Sherman Act should apply in cases of patent holdup that do not involve deception).

54. *See Rambus Inc. v. FTC*, 522 F.3d 456, 466–67 (D.C. Cir. 2008), *cert. denied*, 129 S. Ct. 1318 (2009); *Broadcom Corp. v. Qualcomm Inc.*, 501 F.3d 297, 311 (3d Cir. 2007). In each case, the court held a showing of deception that led to the acquisition of market power was required to state a claim under the Sherman Act.

55. It is not clear that any SSO disallows injunctions. In fact, industry players have argued that, as an example, “European Telecommunications Standards Institute (ETSI) policies do not contain any provision precluding members from seeking injunctive relief when an infringer and potential licensee has rejected a FRAND licensing offer from the patent holder.” *See James Ratliff & Daniel L. Rubinfeld, The Use and Threat of Injunctions in the RAND Context*, 9 J. COMPETITION L. & ECON. 1, 7 (2013). In addition,

Most of the SSOs and their stakeholders that have considered these proposals over the years have determined that there are only a limited number of situations where patent hold-up takes place in the context of standards-setting. The industry has determined that those situations generally are best addressed through bi-lateral negotiation (and, in rare cases, litigation) as opposed to modifying the SSO’s IPR policy and arguably unnecessarily burdening the standardization process for the many ICT standards that are being widely implemented in the marketplace with no apparent IPR-related challenges.

Letter from Microsoft to the Fed. Trade Comm’n, Comment for Patent Standards Workshop, Project No. P11-1204, 3-5 (June 14, 2011), *available at* <http://www.ftc.gov/os/comments/patentstandardsworkshop/00009-60523.pdf>.

56. Statement of the Federal Trade Commission, *Negotiated Data Solutions LLC*, FTC File No. 051-0094 (Jan. 23, 2008), *available at* <http://www.ftc.gov/os/caselist/0510094/080122statement.pdf>.

57. *See, e.g., In re Negotiated Data Solutions, LLC*, FTC File No. 051-0094 (Jan. 23, 2008) (Kovacic, Comm’r, dissenting), *available at* <http://www.ftc.gov/os/caselist/0510094/080122kovacic.pdf>.

58. *See generally Hesse, supra* note 43; *see also* George S. Cary, Mark W. Nelson, Steven J. Kaiser & Alex R. Sistla, *The Case for Antitrust Law to Police the Patent Holdup Problem in Standard Setting*, 77 ANTITRUST L.J. 913 (2011). For an economic and legal defense of the proposition that patent holdup – and in particular patent holdup involving mere breach of contract rather than deception in the standard-setting process – should not be governed by antitrust law, *see* Bruce H. Kobayashi & Joshua D. Wright, *The Limits of Antitrust and Patent Holdup: A Reply to Cary et al.*, 78 ANTITRUST L.J. 701 (2012).

Consumer Welfare In Competition And Intellectual Property Law

BY HERBERT HOVENKAMP¹

Whether antitrust policy should pursue a goal of “general welfare” or “consumer welfare” has been debated for decades. The academic debate is much more varied than the case law, however, which has consistently adopted consumer welfare as a goal.

While some practices such as mergers might produce greater gains in productive efficiency than losses in consumer welfare, identifying such situations would be extraordinarily difficult. First, these efficiencies would have to be “transaction specific,” meaning that they could not be attained by other means. Second, these would necessarily be gains that accrue at lower output levels than previous to the practice; otherwise there would be no consumer harm to balance. But most efficiency gains accrue at higher rather than lower output levels. Third, collusion facilitating practices spread welfare losses across an entire industry, while production gains typically accrue only to the participants in a merger or similar practice. Fourth, the reigning tradeoff models generally assume a market that was competitive prior to the practice and monopolized after. Most practices that facilitate the exercise of market power occur in markets that were noncompetitive to begin with. In these situations consumer losses are relatively larger and producer gains smaller.

Relatively little has been written about consumer welfare and intellectual property law. A well functioning IP system would increase consumer welfare in both the short and long run, so no tradeoff would be needed. In the imperfect system that we have, however, consumer losses do occur, mainly when IP rights are excessive in relation to innovation incentives, or when the rights are given to things that would have been developed (or have already been developed) anyway by ordinary market processes.

Both competition law and intellectual law are concerned with promoting economic welfare. Two fundamental questions for both are determining how welfare should be defined, and how these welfare goals should be implemented. Producer welfare rises as the amount producers receive exceeds the lowest amount they are willing to accept, which is generally their cost. Consumer welfare rises with the difference between the amount consumers must pay and the amount they are willing to pay. When we speak of economic welfare we ordinarily mean the sum of these two differences, or total “surplus.” A perfectly competitive economy is said to be efficient because it maximizes the sum of producers’ and consumers’ surplus. In addition, competition drives selling prices toward marginal cost, and purchasing prices to the marginal consumer’s willingness to pay, assuming there is no price discrimination.

BOTH COMPETITION LAW AND INTELLECTUAL LAW ARE CONCERNED WITH PROMOTING ECONOMIC WELFARE. TWO FUNDAMENTAL QUESTIONS FOR BOTH ARE DETERMINING HOW WELFARE SHOULD BE DEFINED, AND HOW THESE WELFARE GOALS SHOULD BE IMPLEMENTED.

Innovation complicates this model by incorporating incentives over the long run. Firms want to break out of competitive returns by doing something different, whether developing a new product or process, or simply

differentiating their product offering from someone else's. Over the long run innovation contributes a great deal to economic growth, much more than the general movement of markets from lesser to greater price competition.² Private innovation is costly, however, and requires an inducement. The principal inducement the legal system provides is exclusive rights that facilitate short-run returns above competitive levels. The Patent and Copyright Act's temporary periods of exclusivity create exclusive rights that diminish competitive pressures and yield supracompetitive returns for successful inventions.

An ideal innovation policy would facilitate the optimal amount of innovation. Privately, this occurs when the amount of increasing investment in innovation just equals the incremental return.³ In order to be socially optimal, the returns must also net out the social value of the innovation and the inefficiency, or deadweight loss, that results from any exclusive rights that IP protection provides. A precisely tailored policy must determine the optimal duration and scope of intellectual property rights that would produce this result. For example, a right that simply forbids copying, such as copyright and trade secret law convey, is much narrower than a right that condemns all duplication, whether or not the infringer copied or even knew about the right, such as patent law conveys. A patent that lasts 100 years might provide a greater incentive to innovate in some markets, but the longer period would also yield greater deadweight loss by postponing competition.

Identifying and achieving the optimal social level of innovation would be heroic. *Ex post*, one can often conclude that a certain investment in innovation was or was not cost justified. *Ex ante*, however, successful innovation is dominated by the unexpected. Over time, the managers of innovative firms may acquire considerable experience in distinguishing worthwhile from less promising research investments, but the level of uncertainty is almost always higher than when we are considering ongoing production of unchanging products sold in established markets. Further, while the costs of innovation are specific to the industry and the project, the legal system meters returns by providing largely one-size-fits-all regulation of the duration and scope of IP rights. For example, a patent lasts twenty years from the application date, with some adjustments. This applies in both the chemical and pharmaceutical industries, where products can have a commercial life of many decades, and also in electronic communications and computers, where technologies often become obsolete in a few years.

IDENTIFYING AND ACHIEVING THE OPTIMAL SOCIAL LEVEL OF INNOVATION WOULD BE HEROIC. EX POST, ONE CAN OFTEN CONCLUDE THAT A CERTAIN INVESTMENT IN INNOVATION WAS OR WAS NOT COST JUSTIFIED. EX ANTE, HOWEVER, SUCCESSFUL INNOVATION IS DOMINATED BY THE UNEXPECTED.

The transaction costs of operating a system that took individual market differences into account would be very high. Further, we have extraordinarily poor information about such basic questions as the optimal duration and scope of a patent or copyright. Further, in at least some situations the optimal amount is zero.⁴

Competition policy is also concerned with maximizing welfare. A significant debate has developed over whether "general welfare" or "consumer welfare" should be the goal of the antitrust laws. General welfare looks

at the sum of consumers' plus producers' surplus, while consumer welfare looks only at the former. Economists who write about competition policy almost always speak about general welfare. This includes both the Harvard and Chicago Schools, the two groups that have dominated antitrust economics over nearly a century. The debate among antitrust writers is more balanced, with many advocating consumer welfare as antitrust's goal.⁵

The debate offers a degree of richness and complexity that is completely belied by the case law, however.

With virtually no exceptions, the courts take a consumer welfare approach to antitrust and competition law. A recent example is the Supreme Court's divided *Actavis* decision on pay-for-delay settlements in pharmaceutical drug patent cases. The five-member majority evaluated the issue entirely from a consumer welfare perspective. But even the three dissenters acknowledged that "the point of antitrust law is to encourage competitive markets to promote consumer welfare."⁶

To be sure, producer welfare (surplus) is relevant to background policy decisions, such as why we have a rule of reason and when we apply it. But when it comes to specifics the courts uniformly apply a consumer welfare approach.

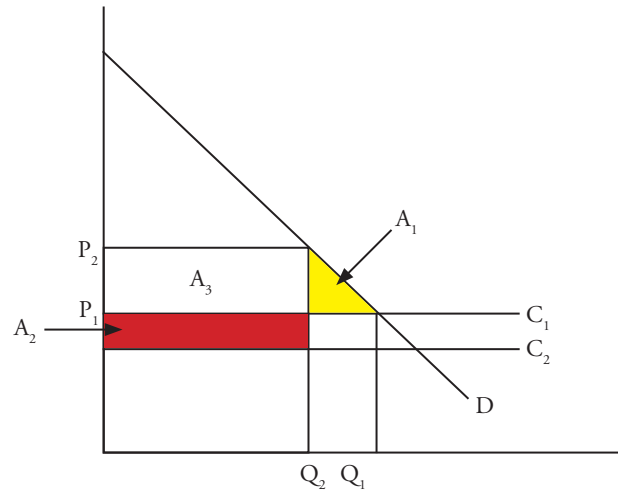
This is not to say that producer gains are taken lightly. On the contrary, they are accorded great weight, but principally because they lead to consumer gains. For example, the market power/structure screen used for rule of reason and merger analysis assumes that consumers can only gain from efficiencies that firms without market power attain.⁷ If 10 Italian restaurants in Manhattan decide to grow tomatoes collectively, competitive harm in the market for either tomatoes or Italian restaurant prices is unlikely. This joint venture will be profitable only if it makes cheaper or better tomatoes, and consumers can only benefit. The same thing would be true if any two of those restaurants should merge. In general, efficiencies from joint activities in competitive markets benefit consumers, but they also benefit producers by giving them advantages over rivals who have not yet duplicated the efficient technology or organization. At the other extreme, naked collusion benefits producers and harms consumers, and antitrust condemns it.⁸

Complexities occur in antitrust efficiency analysis when a practice simultaneously produces efficiency gains and leads to reduced competition, such as some mergers or joint ventures where the participants have significant market power. In general, these situations can be divided into two kinds. First are situations where the firms involved in the challenged activity have or threaten to have serious market power, but the activity produces efficiencies so significant that they fully offset any likely power effects. Prices are lower, or at least no higher, than they had been previously. A good example in the joint venture context is the Supreme Court's *Broadcast Music* (BMI's) decision, which involved copyright blanket license agreements that included virtually every piece of commercially recorded music. The scope of the arrangement created an inference of power, but it was largely undermined by the fact that individual agreements with copyright holders were all nonexclusive, meaning that any one of BMI's 20,000 artists or ASCAP's 22,000 members was free to make unlimited sales outside of the blanket license agreement. Since cartels can profit only by restricting output, achieving that result in this situation was unthinkable.⁹ Further, the licensing arrangement, which created the play-on-demand authorization that programmers use to this day, was so efficient that the market as we know it could not exist. So there was no balancing of consumer losses against producer gains. Both groups were winners, leaving nothing to balance.

Another example is the merger that both creates monopoly power or facilitates its exercise, and also yields production efficiencies so substantial that the post-merger price (or quality) is better for consumers than the pre-merger situation had been. Once again, there is nothing to trade off. The merging firms may have higher price/cost margins, profiting them, but consumers are better off as well. The merger should be lawful under either a general welfare or a consumer welfare test. This is in fact the test that the Government applies under its Merger Guidelines. In order to defend a *prima facie* anticompetitive merger on the basis of efficiencies the proponents must show that the efficiencies are "merger specific"—that is, that they cannot readily be achieved

without the merger. They must also show that proven efficiencies are of sufficient magnitude that the market price following the merger will be no higher than it had been before.¹⁰ In that case there is nothing to trade off.

The more difficult cases occur when a tradeoff is necessary. In the 1960s, Oliver E. Williamson famously illustrated that a merger that actually raised prices could nonetheless improve total welfare if the production efficiency gains resulting from the merger were greater than the consumer losses. He also concluded that relatively modest efficiency gains could achieve this goal.¹¹



The figure shows the result of a merger, joint venture or other practice that both creates monopoly and produces gains in productive efficiency. Prior to the practice the market was competitive. Prices were at P_1 , which is equal to cost C_1 . The challenged practice causes two things. First, the firm's costs decline to C_2 . Second, however, the firm acquires a monopoly and no longer prices at cost. Rather it reduces output to Q_2 and raises price to P_2 .

Under this model, which Williamson conceded to be highly simplified, consumer prices go up, producing a monopoly deadweight loss equal to A_1 in the figure. However, productive efficiency also goes up, producing gains of A_2 , which in this particular figure are clearly larger than the A_1 losses. So the practice is efficient in the economic sense of increasing total welfare, even though prices are higher. While A_3 also represents producer gains and consumer losses, it is a “wash” because producers are better off by the same amount as consumers are worse off.

Williamson's model has been subjected to a fair amount of criticism. First, if the merger facilitates market wide collusion, which is a common rationale for condemning mergers, then the price increases will occur across the market, but only the merging firms will enjoy the production cost savings. This could change the calculus considerably in collusion-facilitating mergers of, say, five-to-four players, or four-to-three players.

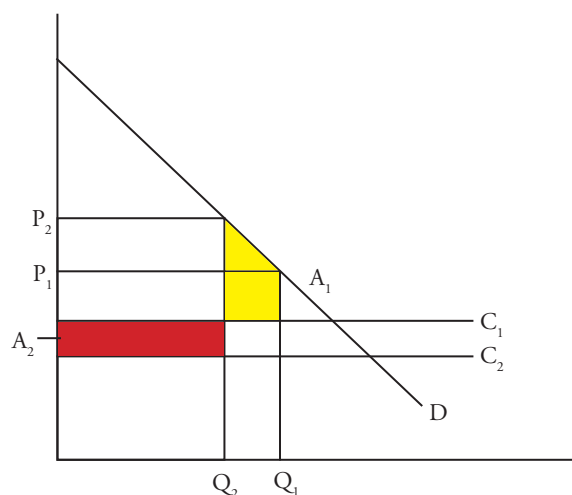
Second, the analysis assumes that the efficiencies are strictly “merger specific,” which means that only the merger that reduces competition can produce them. Often efficiencies can be attained in less harmful ways, including licensing as an alternative to acquisition, or partial spinoffs to other sellers.

Third, the efficiencies that Williamson's model illustrates usually have to come from some effect other than scale economies, because in the tradeoff situation output is actually lower than it had been before. To be sure,

some mergers can enable firms to take advantage of scale economies even as they reduce output. An example might be two firms that each have an inefficient 10,000 unit plant. After a merger the firm might produce 18,000 units efficiently out of a single larger plant. Note that the merger itself does not achieve this result, however. It simply leaves the firm with two inefficiently small plants. None of this is to suggest that there are no efficiencies that can be attained at reduced output; however, the universe of efficiencies that occur at lower rather than higher output may be small, particularly when one adds in the requirement that they must be merger specific.

Fourth, some care must be taken to ensure that the cost savings are not merely pecuniary. For example, a merger that creates a monopoly on the selling side might also create significant power on the buying side, enabling the firm to suppress the prices that it pays. But in that case any gains to the merging firm could be more than offset by the losses that accrue to its suppliers, and the case for overall efficiency evaporates.

Fifth, the Williamson model assumes a market that was perfectly competitive before the challenged practice but monopolized afterward. This would be a rarity. More likely the market prior to the merger or joint venture was already noncompetitive, but to a lesser degree. One of the reasons that Williamson's picture shows a small deadweight loss is because at the competitive level the sales are taken from "marginal" consumers who place a low value on the product. As a result, loss of these sales entails a fairly low deadweight loss. At higher levels the amount of surplus per consumer is much greater, making an output reduction of the same magnitude more costly. Further, at these higher levels any efficiency gains must be spread over a lower output.¹²



The second figure illustrates this idea. It shows the same market as the first figure, and with a merger or other practice that produces the same per unit cost reduction. In this case, however, the market was already noncompetitive to begin with, reflecting $P1$ prices that were higher than cost $C1$. As a result two things happen. First, consumer deadweight loss is larger because output is being taken from consumers whose willingness to pay is higher in relation to the product's cost. Second, because output is already lower to begin with, the efficiency gains resulting from a further output reduction are spread over a smaller number of units. Even though the demand curve is identical to the one in the first figure and the per unit amount of the efficiency gains (the height of the rectangle $A2$) is the same, it is now no longer clear that the area covered by the red figure is greater than the area of deadweight loss defined by $A1$.

Finally, note that the "tradeoff" model applies to a limited range of situations—namely, where the ef-

efficiency gain is sufficiently large to offset the deadweight loss, but not so large that it actually results in a lower price than before the merger or joint venture occurred. If the gains were that large there would be no tradeoff; both consumers and producers would be better off. It is difficult to say what percentage of mergers, joint ventures or other practices lead to both increased market power and a higher price, but offset production gains that exceed consumer losses.

One reason that the consumer welfare principle entirely dominates antitrust analysis is administrability. Anytime measurement makes a difference—that is, where a practice both facilitates an actual output reduction resulting from market power but also produces efficiencies—the fact finder would have to quantify these effects and net them out. That would require a cardinal measurement of deadweight loss and offsetting gains in production efficiency.

Measuring deadweight loss is much different from measuring simple consumer overcharges. Deadweight loss is equal to the amount of consumers' and producers' surplus that is lost as the result of a monopoly output reduction. Computing this requires knowledge about the demand and supply curves in a region where there is currently no output. Quantifying efficiency gains would require the fact finder to identify merger specific gains and then compute the value of either cost savings or product or service improvement over the post-practice output. I know of no court that has even attempted this in a situation where both numbers are positive and significant. Rather, they cite efficiencies as justifying a practice when no market power is present or it is clear that the practice is not reducing output at all. Alternatively, they cite competitive harm in situations where market power effects indicate an actual output reduction, ignoring efficiencies. The approaches taken in these two situations are tractable because there is nothing to balance.

ONE REASON THAT THE CONSUMER WELFARE PRINCIPLE ENTIRELY DOMINATES ANTITRUST ANALYSIS IS ADMINISTRABILITY. ANYTIME MEASUREMENT MAKES A DIFFERENCE—THAT IS, WHERE A PRACTICE BOTH FACILITATES AN ACTUAL OUTPUT REDUCTION RESULTING FROM MARKET POWER BUT ALSO PRODUCES EFFICIENCIES—THE FACT FINDER WOULD HAVE TO QUANTIFY THESE EFFECTS AND NET THEM OUT. THAT WOULD REQUIRE A CARDINAL MEASUREMENT OF DEADWEIGHT LOSS AND OFFSETTING GAINS IN PRODUCTION EFFICIENCY.

Setting aside these administrative difficulties, one historical defense of a total welfare approach is that producers compete with one another and will compete away any producer gains into consumer gains.¹³ Of course, if we have calculated correctly, then the market power created by a practice will be sufficiently durable that we cannot have confidence that this will happen in the near term. More fundamentally, however, there is no reason for giving a preference to producer competition over consumer robustness. To the extent a practice makes a market bigger there will be more demand and more inducement to innovate. *A priori*, increased output and even efficiency is just as likely to “trickle up” from increased consumer demand as it is to trickle down from decreased producer supply.

Suppose that we have identified an efficiency that can be attained only by an output-reducing merger.

What will be the impact of a legal rule that prohibits this merger as long as the output reduction is likely? First, the firms might simply give up their efforts to attain this particular efficiency. Second, they might try to fix the merger in such a way as to preserve the efficiency but blunt the competitive harm; as a result, output will be

higher than previously, or at least not any lower. Third, they might try to attain the efficiency by some other means than merger. For example, a large firm that wants a smaller firm's technology might be able to get it with a nonexclusive license, which facilitates the efficiency but without the competitive harm. Fourth, they might attempt to increase the magnitude of any efficiencies so that resulting prices will be no higher than they were before, thus eliminating any need for trading off. Any one of these alternatives would make everyone better off. Given our inability actually to balance consumer harm from output reductions against productive efficiency gains, antitrust's insistence that a practice do no harm to consumers makes a good deal of sense.

Nevertheless, the importance of Williamson's conclusion remains: a significant set of mergers (or other practices, such as joint ventures) could exist in which the practice causes an actual market-power-induced price increase, but offsetting efficiencies make overall effects positive under a general welfare test. The window is narrow, however. The efficiencies must be large enough to offset deadweight loss, but not so large as to offset the consumer overcharge altogether.

When one looks at actual antitrust policy, the set of cases trading off actual productive efficiency gains against actual consumer losses is close to empty. Courts simply do not find, first, that a practice actually facilitates the exercise of market power and leads to higher prices, but second, that the practice should be approved because producer gains are bigger than consumer losses. Instead, while efficiency gains are important and often even central, the court must be assured that there are no consumer losses at all—either because market power cannot reasonably be exercised or else because the efficiency gains are so significant that they fully offset any market power effects.

What happens to the consumer welfare principle when we think about innovation and intellectual property policy? At first glance IP law seems different. Courts and other writers often speak of patent and copyright interests as “monopolies.” Inherent in the concept of exclusive IP rights is that they provide returns above the competitive level for a period long enough that developers can recoup their research and development costs.

If the IP system does what it should be doing, however, consumer loss is not an inherent part of the design. New innovation typically competes with existing technology, which largely stays in place. Consumers obtain the benefit of the new products or processes that IP rights make possible, and these should yield a surplus even at monopoly prices.

WHEN ONE LOOKS AT ACTUAL ANTITRUST POLICY, THE SET OF CASES TRADING OFF ACTUAL PRODUCTIVE EFFICIENCY GAINS AGAINST ACTUAL CONSUMER LOSSES IS CLOSE TO EMPTY. COURTS SIMPLY DO NOT FIND, FIRST, THAT A PRACTICE ACTUALLY FACILITATES THE EXERCISE OF MARKET POWER AND LEADS TO HIGHER PRICES, BUT SECOND, THAT THE PRACTICE SHOULD BE APPROVED BECAUSE PRODUCER GAINS ARE BIGGER THAN CONSUMER LOSSES. INSTEAD, WHILE EFFICIENCY GAINS ARE IMPORTANT AND OFTEN EVEN CENTRAL, THE COURT MUST BE ASSURED THAT THERE ARE NO CONSUMER LOSSES AT ALL—EITHER BECAUSE MARKET POWER CANNOT REASONABLY BE EXERCISED OR ELSE BECAUSE THE EFFICIENCY GAINS ARE SO SIGNIFICANT THAT THEY FULLY OFFSET ANY MARKET POWER EFFECTS.

Consider the world just before Pfizer's blockbuster drug Lipitor, which reduces LDL cholesterol, was brought to market. When Lipitor was not yet available it did not create a consumers' surplus, and consumers used whatever cholesterol-reducing products and activities that the market offered. The drug was then introduced at a very high price. Consumers were not injured by this transition, however, whether or not the price was set at monopoly levels. People who did not buy Lipitor were unaffected. Not only were the earlier alternatives still available, they may even have become cheaper by virtue of Lipitor's competition. By contrast, those who did buy Lipitor obtained at least as much value as its price. The fact that it is purchased at all entails that it is producing consumer wealth. Total output of cholesterol reducing formulations has gone up rather than down. To be sure, Lipitor might not be creating as much consumer wealth as it would if it were competitively priced, but it is clearly creating more consumer wealth than a market without Lipitor or its equivalent. The relevant comparison is not between competitively and monopoly priced Lipitor, but between monopoly-priced Lipitor and a market with no Lipitor at all.

A subset of patents operates differently in that the patents actually remove older consumer choices from the market. For example, a cost reducing process patent might give its owner so large an advantage over rivals that it is able to drive them out of business. Further, because consumers don't get to select the process, the older process might be unavailable as an alternative. Even here, however, the patentee would not be able to charge more for the product than it did before its new process was developed, because the older technology is still profitable at the pre-invention price. For example, suppose that under the older technology widgets could be produced for \$5 each. The patentee invents a technology that enables production of equally good widgets at \$4. As long as the patentee held its output price under \$5 it would be in a position to drive out rivals using the old technology. However, if it attempted to use its newly acquired monopoly position to set a price higher than \$5, the old technology would remain profitable. Consumers would not be worse off, and they could be much better off.

The all-important "welfare" question for legal policy is the role of the patent system in getting socially beneficial products and processes developed. Consumer harm does result when the intellectual property system provides more exclusionary power than is necessary to develop some new thing, or when it excludes without providing anything new at all. For example, if a 10-year patent rather than the actual twenty year patent was all that was needed to motivate the development of Lipitor, then the additional patent coverage is both a wealth transfer from consumers and a deadweight loss.

Consumer harm also results when the patent system permits a firm to obtain exclusive rights on something that is either publicly available or that soon would be in the ordinary course of events. That is hardly the case with Lipitor, a drug that was in very high demand after it was introduced, and that led to robust generic competition when the Lipitor patent expired late in 2011. In general, pharmaceutical patents on active ingredients are among the most robust patents in the system, and patent coverage is warranted for the soundest of reasons: development costs are high but copying costs are typically low. By contrast, "evergreened" extension patents on drugs, which are typically on new dosages or delivery mechanisms rather than the active ingredient, have a significantly higher failure rate and, when granted, can harm consumers by extending patent exclusivity periods.¹⁴

Patents are sometimes granted on ideas that do not meet patent law's requirement of nonobvious subject matter. These ideas would soon have been developed, or in some cases they have already been developed in areas that are not readily searchable. Certain industries, such as information technologies, or certain types of patents, such as software and business method patents, are particularly prone to such abuses. "Obvious" patents harm

consumers by creating exclusive rights in things that would otherwise sell competitively. The deadweight loss of obvious patents is high. As the previous discussion suggests, it tends to be higher in markets that are already noncompetitive to begin with, because marginal consumers in those markets have more consumers' surplus to lose. Many pharmaceutical extension patents very likely fall into this category.

The deadweight loss caused by obvious patents results partly from an overly generous patent-granting process. But we would have to put far more resources into patent examination and prosecution than we currently do in order to catch most problems effectively.¹⁵ One thing that Congress could and should change is the statutory presumption of validity. The presumption is something of an oddball. Most presumptions are recognized because the association of two things is highly probable. For example, a letter that has been mailed is presumed to have been delivered, because nearly all are.¹⁶ In that case it makes sense to assign the burden of proof to the person with the least plausible claim. But litigated patents are found to be invalid anywhere from one-third to one-half of the time, even under the current presumption.¹⁷ Given that, the patentee should have the burden of proving its validity case, just as it must prove infringement and damages.¹⁸ Nevertheless, Congress is unlikely to change the statutory presumption, and in 2011 the Supreme Court re-affirmed that it can be defeated only by clear and convincing evidence.¹⁹

More generally, the problem of socially harmful patents lies with a statutory drafting process in both patent and copyright law that has persistently placed producers in charge, paying little attention to consumer interests. That was true of the 1952 Patent Act, and even truer of the 1976 Copyright Act.²⁰ More recently the America Invents Act began as a series of high minded and excellent ideas but the drafting was quickly taken over by a veritable war among various producer groups. Consumer interests were largely ignored.²¹

...LITIGATED PATENTS ARE FOUND TO BE INVALID ANYWHERE FROM ONE-THIRD TO ONE-HALF OF THE TIME, EVEN UNDER THE CURRENT PRESUMPTION.¹⁷

Permitting producers to control the law making process is sad because in this particular case consumers have the right set of incentives to produce optimal innovation policy. By and large consumer interests favor low prices, high output, high quality, and cost justified improvements.

Producer interests are much less aligned with the public interest, although their situation is far more complex. First, producers generally favor high margins on their own output, although they also want lower prices on inputs that they purchase from others. Second, to the extent producers are owners of intellectual property rights and profit either by excluding or licensing, they tend to regard more as better. They want longer terms and broader scope. By contrast, if they appear on the market as both owners (licensors) and licensees, their interests become more complex. If technology is fast moving and has a fairly short commercial life, as it is in many information technologies, then first mover advantages are more significant protectors of innovation returns. Patenting may be little more than a socially costly nuisance. In general, producers want their own patent portfolios to be as valuable as possible and those of rivals or others in a position to license to them to have as little value as possible. In many patent-rich environments firms that do a great deal of patenting profit from ambiguity and indeterminacy in the system. This hotchpot of producer interests is hardly calculated to produce an optimal system.

To be sure, some producers are aligned more closely with consumers. These tend to be firms that do not do a

great deal of protectable innovation themselves but who purchase their IP-protected inputs from other suppliers, and whose own IP portfolios are minimal—typically some trademarks and perhaps a few copyrights or trade secrets. A great deal of small business falls into this category, particularly retailing, as well as a fair amount of traditional manufacturing in commodity industries. But these groups have not been well represented in the patent legislation process either.

One well-known example of excessive producer orientation is the ever increasing length of copyright protection, even as the commercial shelf life of most copyrighted products is becoming shorter. If the incentive to produce requires a measured period of protection during a copyrighted product's commercial life, it follows that no protection should be required once the commercial life has come to an end. In fact, the vast majority of copyrighted works being produced today have a commercial life that falls far, far short of the century or so of

IN FACT, THE VAST MAJORITY OF COPYRIGHTED WORKS BEING PRODUCED TODAY HAVE A COMMERCIAL LIFE THAT FALLS FAR, FAR SHORT OF THE CENTURY OR SO OF PROTECTION THAT THE COPYRIGHT ACT OFFERS.²² INDEED, FOR A SIGNIFICANT PORTION OF COPYRIGHTED WORKS, THE STATUTE DOES PRECISELY THE OPPOSITE OF WHAT IT SHOULD BE DOING: IT HAS MADE THE WORKS VIRTUALLY UNAVAILABLE.

protection that the Copyright Act offers.²² Indeed, for a significant portion of copyrighted works, the statute does precisely the opposite of what it should be doing: it has made the works virtually unavailable. For example, out-of-print books published in 1923 or later may still be under copyright, but a researcher can obtain them only by going to a library or a used bookstore, which may or may not have them. The same thing is often true for journal articles, software, and at least a fair amount of music, photography and video content. Authors and assignees are getting either no or trivial royalties on these out-of-print works, but the public is not getting access to them either. One solution for this problem is a copyright act that is more aligned with our empirical knowledge about the changing value of copyright works over time. For example, a statute that offered a

short period of protection accompanied by a moderately costly right of periodic renewal would serve to keep things under protection as long as the copyright holder reasonably expected a positive commercial return.²³

Unfortunately, the consumer interests that are best aligned with optimal innovation are also the interests that have the smallest amount of involvement in the IP systems themselves. In the copyright system consumers can be sued, although most of the lawsuits are limited to illegal downloading and pirating. Even in copyright, however, producers have largely controlled legislative processes. Patent infringement lawsuits against consumers are infrequent, and as a result the patent process is largely hidden from them. Most patent activity, even in patent-rich environments such as cell phones or computers, occurs upstream in the distribution chain. Excessive patent coverage or litigation costs are certainly reflected in higher consumer prices, but to most consumers the cause is invisible.

Quite naturally, when Congress seeks advice on how to improve patenting it goes to those who have the experience and expertise. This is a common occurrence and explains a great deal of regulatory capture. For example, those who generate electricity and must procure the technologies, build the networks, and determine the fuels know a great deal more about power generation than someone whose principal experience is turning on a light switch. For legislative or regulatory bodies to go to producers rather than consumers for regulatory

advice is hardly an irrational thing. But this makes the regulatory capture phenomenon much more fundamental to regulation than many people realize. Patent law is particularly prone to this. Not only is the law old, idiosyncratic, complex, and technical, but consumers are hardly involved in the process at all except indirectly as purchasers or users of patented technology. It is only natural for a legislative body or agency to go to those with the experience.

One important step toward greater consumer orientation would be development of a “political economy” of the patenting process, aligning patent doctrine more closely with economic or innovation effects. Here, antitrust has a distinct institutional advantage over patent law. Through decades of study, antitrust enforcers have developed empirically-supported models predicting how practices such as mergers affect pricing and output, the links between industry structure and propensity to collude, the values and costs of vertical integration and contract practices, and the like. One should not push the point too far, because empirical study has hardly eliminated controversy. Nevertheless, at a fundamental level, antitrust policy making today has an empirical content that is rooted in economic theory about how markets work.

Notwithstanding its strong identification with newness and innovation, patent law is in fact built on an ancient appropriation model borrowed from the law of property, particularly land titles, in which the economic effects of particular technical rules were seldom made explicit. In this sense patent law resembles property law before the rise of cost-benefit analysis approaches to public land use policy and environmental law. While economic studies of innovation are manifold and have produced a rich literature relating innovation rates to such things as market structure,²⁴ we know surprisingly little about the effects of specific patent doctrines.

As a result, patent law does a much poorer job of “metering” innovation than antitrust law does of metering competition, deficiencies notwithstanding.

One example of this is patent law’s nonobvious requirement, which differs from the novelty requirement and is much more difficult to apply. Novelty fails when something in the prior art reveals that the thing for which the patent is claimed already exists, a backward looking question asking whether something in the prior art anticipates all of the elements of a patent claim.²⁵ By contrast, non-obviousness, or “inventive step” in European law, considers whether a person skilled in this particular art would be likely to have come up with this idea on her own.²⁶

Determining nonobvious subject matter requires going beyond what the prior art actually contains in order to assess whether someone of “ordinary” skill, and who is acquainted with the prior art in that field would be

FOR LEGISLATIVE OR REGULATORY BODIES TO GO TO PRODUCERS RATHER THAN CONSUMERS FOR REGULATORY ADVICE IS HARDLY AN IRRATIONAL THING. BUT THIS MAKES THE REGULATORY CAPTURE PHENOMENON MUCH MORE FUNDAMENTAL TO REGULATION THAN MANY PEOPLE REALIZE. PATENT LAW IS PARTICULARLY PRONE TO THIS. NOT ONLY IS THE LAW OLD, IDIOSYNCRATIC, COMPLEX, AND TECHNICAL, BUT CONSUMERS ARE HARDLY INVOLVED IN THE PROCESS AT ALL EXCEPT INDIRECTLY AS PURCHASERS OR USERS OF PATENTED TECHNOLOGY. IT IS ONLY NATURAL FOR A LEGISLATIVE BODY OR AGENCY TO GO TO THOSE WITH THE EXPERIENCE.

likely to come up with the invention independently. In the patent granting process non-obviousness queries typically involve situations where there are multiple pieces of prior art but no single piece fully anticipates a particular patent claim; or where the invention is anticipated in a different market, or “field of endeavor,” but not in the one where the patent is sought. In addition, the non-obviousness inquiry may involve considerations that do not show up in the patent prior art, such as acknowledged but unmet needs, or general experience and

EVEN WHEN REMEDIES ON UNPRACTICED PATENTS ARE LIMITED TO DAMAGES, MANY PATENTS HAVE A LITIGATION VALUE IN EXCESS OF THEIR TECHNOLOGICAL VALUE, WHICH MAY IN FACT BE ZERO. AS A RESULT, LARGE-SCALE LITIGATION BY PATENT AGGREGATORS IS ACTUALLY CREATING A MARKET FOR BAD PATENTS.³⁵

understanding in an industry.²⁷ For example, evidence that others skilled in the art had tried to solve a problem but failed points toward nonobvious subject matter.²⁸ Examiners may also rely on their own expertise in a field to intuit whether the claimed invention is nonobvious. The Supreme Court has observed that in many fields there may be “little discussion of obvious techniques or combinations, and it often may be the case that market demand, rather than scientific literature, will drive design trends.”²⁹ The Patent Office’s examination guidelines acknowledge that the printed

prior art, both patent and non-patent, will not invariably answer every question.³⁰

Neither the case law on nonobvious subject matter nor the patent examination guidelines suggest empirical studies of the extent to which accused infringers are copyists rather than independent inventors. Patent examiners have neither the time nor the resources to conduct such studies in determining whether to grant a particular patent. But the forward-looking question of ambiguity can never be addressed effectively by backward looking inquiries revealing what the prior art contains. The real thing we want to know is whether the inventor is contributing something that is worthy of an exclusive right because society was unlikely to get this particular invention through ordinary market processes.

Here, evidence about the extent of independent invention in different areas could be very helpful. The popular conception of a patent infringer may be the person who willfully copies someone else’s invention. In fact, the patent infringement case law reveals relatively few copyists. Only some 11 percent of complaints allege “willful” infringement and only 30 percent allege that a defendant actually knew about the patent it was infringing. Further, only 2 percent percent of reported decisions include a fact finding of actual copying.³¹ This is all the more important because the Patent Act provides heightened damages for willful infringement.³² If patentees knew that alleged infringers had copied their patented technology they would have every motive to bring willful infringement claims. Further, the copying numbers vary significantly by industry, ranging from a high of 20 percent in pharmaceuticals, where the value of copying is obvious, to very close to zero in biotechnology, semiconductors, electronics, and computers.³³ This suggests that in at least some markets numerous patents are being granted for technologies that were likely to be produced through ordinary competitive processes.

To be sure, distinguishing copying from independent invention is not necessarily easy. Someone who buys one unit of a patented product may be able to identify its patented technology without leaving much of a record. However, many patents relate to processes rather than products and are not readily discoverable. Many other patents are not practiced at all, and the only record of their existence is in the patent databases. An unpracticed patent is also creating no surplus: its only “value” is the limitation it places on others to develop and use the covered invention. As a result, granting enforcement of unused patents against independent discoverers of the

patent technology is perverse. It actually removes technologies from consumer availability, precisely the opposite of what the patent system should be contemplating. Subsequent to the Supreme Court's *eBay* decision the lower courts have increasingly been limiting remedies on unpracticed patents to damages.³⁴ But that is hardly a complete solution. Even when remedies on unpracticed patents are limited to damages, many patents have a litigation value in excess of their technological value, which may in fact be zero. As a result, large-scale litigation by patent aggregators is actually creating a market for bad patents.³⁵

We have never developed a strong empirical literature on rates of independent discovery and how it varies from one technology to another. The "property" queries that the courts and examiners currently use are not well designed for determining the extent of independent invention. The evidence would certainly be market specific. For example, in the pharmaceutical industry it is highly unlikely that a firm would duplicate a patented molecule unless it had knowledge of the patent. By contrast, business method patents are routinely given on trivial changes in distribution that others often come up with on their own. Copying of minor advances is the way that competition works in many industries. The real impact of excessive business method patenting is to create unwarranted product differentiation that reduces market output while providing little or nothing in return.

Closely related is the problem of patent notice. The obligation to obtain information and the obligation to provide it are correlative. Optimizing requires that the obligation be given to the person who can act at the lowest cost. In general, the cost of providing notice is lower than the cost of searching.³⁶ Thus for example, the real property recording acts place the obligation to record on the landowner, who usually records once. Otherwise the prospective purchaser must do a lengthy and difficult history in order to determine land ownership. "Notice" in patent law requires two different things. First, one must be able to discover relevant patents; second, one must be able to interpret them once discovered. The problems are at least partially interrelated: if a patent is very difficult to interpret it will also be difficult to discover.

The economics of notice suggest that patentees should have the primary duty to provide realistic and timely notice to likely infringers. Of course, issued patents are searchable public records. That does not solve the problem, however, when an area of enterprise has thousands of patents that are cut very thin and notable for their ambiguity. In the context of standard setting the problem is beginning to be addressed through arrangements that require patentees who wish their patents to be declared standards essential to disclose them up front and promise to make them available on FRAND terms. Upon inspection, even standard-essential patents turn out frequently not to be worth much.³⁷

Ambiguity undermines notice when a lengthy, costly interpretation of a patent is necessary before one can determine its coverage and the resulting expert opinions are frequently unreliable. Patent practice encourages ambiguity in many technologies by enabling drafters to have it both ways, particularly for subsequently developed technology. Ambiguity permits a broader construction if no objections are raised, but a narrower construction if the patent encounters problems on anticipation or non-obviousness grounds, or perhaps when the applicant cannot show that it was in "possession" of the invention as broadly construed, which is another way of saying that he did not have a complete conception of it. As Judge Plager wrote in a recent Federal Circuit decision:

Sometimes such ambiguity is the result of sloppy drafting, and sometimes it appears that claims are drafted with a degree of indefiniteness so as to leave room to later argue for a broad interpretation designed to capture later-developed competition. The problem is exacerbated when,

as here, there is a conflicting or indeterminate written description and prosecution history with regard to the claim terms at issue. Claim construction then becomes a game of crystal ball gazing, not resolved until this court's gaze is announced.³⁸

Judge Plager then proposed that patent claim construction follow the general common law rule that ambiguous claims should be construed against the drafter.³⁹ That rule might require some adjustment for patents that have already been issued, but applied prospectively it would lead to clearer and more searchable patents. Indeed, the common law's one bite at the apple rule has always worked quite well for privately drafted documents of all sorts. Ambiguities are construed against the drafter and most of the time we do not permit adjustments after the fact. That gives the drafter a strong incentive for clarity the first time around.

CONCLUSION

Giving recognition to consumer interests in intellectual property will require more than tweaking, although some tweaks could help. Further, because actually

GIVING RECOGNITION TO CONSUMER INTERESTS IN INTELLECTUAL PROPERTY WILL REQUIRE MORE THAN TWEAKING, ALTHOUGH SOME TWEAKS COULD HELP. FURTHER, BECAUSE ACTUALLY INCLUDING CONSUMERS IN THE LAW MAKING PROCESS MAY BE IMPRACTICAL, CONSUMER WISHES MAY HAVE TO BE INFERRED OBJECTIVELY FROM INFORMATION ABOUT PERCEIVED VALUE.⁴⁰

including consumers in the law making process may be impractical, consumer wishes may have to be inferred objectively from information about perceived value.⁴⁰ In any event, rewriting IP law so as to serve consumers would require upending a good deal of history, producing a policy that is driven less by historical property rules and more by empirical economic study linking the benefits and costs of innovation to specific IP doctrines. The courts can certainly do some of the work. For example, the 1970s revolution that moved antitrust law away from small business protectionism was very largely a creature of the courts.⁴¹ Mainly, however, reform would require a special effort by Congress to give properly identified consumer interests or their effective surrogate a seat at the legislative drafting table. ▲

1. Ben V. & Dorothy Willie Professor of Law and History, University of Iowa.
2. The substantial literature is discussed in CHRISTINA BOHANNAN & HERBERT HOVENKAMP, CREATION WITHOUT RESTRAINT: PROMOTING LIBERTY AND RIVALRY IN INNOVATION 7-11 (2012)
3. The American philosopher of science Charles Sanders Peirce made this observation already in the nineteenth century. See Charles S. Peirce, *Note on the Theory of the Economy of Research*, 1875 COAST SURVEY REPORT 197–201, reprinted in THE WRITINGS OF CHARLES S. PEIRCE: 1879–1884 at 73 (Christian J.W. Kloesel, ed., Bloomington: Indiana Univ. Press, 1989).
4. Dan L. Burk & Mark A. Lemley, Policy Levers in Patent Law, 89 Va. L. Rev. 1575 (2003); BOHANNAN AND HOVENKAMP, CREATION WITHOUT RESTRAINT 44-46.
5. See Herbert Hovenkamp, *Implementing Antitrust's Welfare Goals*, 81 FORDHAM L.REV. 2471 (2013) (discussing the alternatives); see generally Barak Orbach, *Forward: Symposium on The Goals of Antitrust*, 81 FORDHAM L.REV. 2451 (2013).
6. FTC Actavis, Inc. 133 S.Ct. 2223, 2238 (2013) (Roberts, C.J., dissenting, joined by Justices Scalia and Thomas; Justice Alito did not participate).
7. See 2B PHILLIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW ¶¶500 et seq (4th ed. 2014) (market power generally); 4 id. ¶901 (3d ed. 2009) (mergers).
8. 12 id. ¶2004 (3d ed. 2012).

9. Broadcast Music, Inc. v. Columbia Broadcasting Sys., Inc., 441 U.S. 1, 5, 11 (1979) (Members “retain the rights individually to license public performances, along with the rights to license the use of their compositions for other purposes”).
10. US Department of Justice, Antitrust Division, and Federal Trade Commission, Horizontal Merger Guidelines §10 (2010, available at <http://www.justice.gov/atr/public/guidelines/hmg-2010.html#10>).
11. Oliver E. Williamson, *Economies as an Antitrust Defense: The Welfare Tradeoffs*, 58 AM.ECON.REV. 18, 21 (1968).
12. See Areeda Hovenkamp ¶970e (4th ed. 2014). Williamson also acknowledged the point: Oliver E. Williamson, *Economies as an Antitrust Defense Revisited*, 125 Univ. Penn.L.Rev. 699, 710 (1977).
13. See ROBERT H. BORK, *THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF* (1978).
14. See, e.g. Ely Lilly & Co. v. Sicor Pharma., Inc., 426 Fed. Appx. 892 (Fed. Cir. 2011) (pharmaceutical extension patent invalid for obviousness-type double patenting); Anthony M. Insogna, et. al, *Obviousness-Type Double Patenting*, 5 LANDSLIDE 37 (No. 1 2012).
15. Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. UNIV.L.REV. 1495 (2001).
16. E.g., Echavarría v. Nat. Grant Mut. Ins. Co., 275 Conn. 4008 (2005); JOHN HENRY WIGMORE, *EVIDENCE* §2519 (Chadbourn ed. 1981 & Supp.).
17. WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 338 (2003).
18. Douglas G. Lichtman and Mark A. Lemley, “Presume Nothing: Rethinking Patent Law’s Presumption of Validity”, in *COMPETITION POLICY AND PATENT LAW UNDER UNCERTAINTY: REGULATING INNOVATION* (Geoffrey A. Manne & Joshua D. Wright, eds., 2011).
19. 35 U.S.C. § 282. See *Microsoft Corp. v. i4i Ltd. Partnership*, 131 S.Ct. 2238 (2011).
20. On patents, see HERBERT HOVENKAMP, *THE OPENING OF AMERICAN LAW: NEOCLASSICAL LEGAL THOUGHT, 1870-1970*, ch. 10 (2014); on copyright, see BOHANNAN & HOVENKAMP, *CREATION WITHOUT RESTRAINT*, *supra* 137-140; William Patry, *Copyright and the Legislative Process: A Personal Perspective*, 14 CARDOZO ARTS & ENT. L.J. 139, 141 (1996).
21. On the interest groups that drove the America Invents Act, see Zach Carter, *The Spoilsmen: How Congress Corrupted Patent Reform*, HUFFINGTON POST (Oct. 4, 2011, 6:12 AM), http://www.huffingtonpost.com/2011/08/04/patent-reform-congress_n_906278.html; Steve Lohr, *Inventor Challenges and a Sweeping Revision in Patent Law*, N.Y. TIMES (Aug. 26, 2012), <http://www.nytimes.com/2012/08/27/technology/mark-stadnyk-challenges-sweeping-revision-in-patent-law.html>; Amanda Becker, *Patent Reform Measure Ignited Fierce Lobbying Effort*, WASH. POST (Mar. 27, 2011), http://articles.washingtonpost.com/2011-03-27/business/35208767_1_first-to-file-system-patent-system-patent-office. See also Charles E. Miller & Daniel P. Archibald, *Beware the Suppression of District-Court Jurisdiction of Administrative Decisions in Patent-Validity Challenges Under the America Invents Act: A Critical Analysis of a Legislative Black Swan in an Age of Preconceived Notions and Special-Interest Lobbying*, J. PAT. & TRADEMARK OFF. SOC’Y 124, 135 (2013); David Orozco, *Administrative Patent Levers*, 117 PENN. ST. L. REV. 1, 49–50 (2012); Renee Kaswan, David Boundy & Ron Katznelson, *Patent Reform: Effects on Medical Innovation Businesses*, PAT. & INNOVATION ECON., Summer 2010, at 11–12.
22. The basic protection is seventy years plus life of author. 17 U.S.C. §302(a). See *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) (upholding Congressional retroactive extension of copyright term).
23. See BOHANNAN & HOVENKAMP, *CREATION WITHOUT RESTRAINT*, *SUPRA*, 400-401.
24. See *id.*, 8-11; Wesley Cohen & Richard Levin, “Empirical Studies of Innovation and Market Structure,” in 2 *HANDBOOK OF INDUSTRIAL ORGANIZATION* 1059–1107 (Richard Schmalensee & Robert Willig eds., 1989).
25. See 35 U.S.C. §102(a), (d), (e).
26. 35 U.S.C. §103.
27. *KSR Intern. Co. v. Teleflex, Inc.*, 550 U.S. 398, 420-421 (2007).
28. *Graham v. John Deere Co.*, 383 U.S. 1, 17-18 (1966).
29. *KSR*, 550 U.S. at 419.
30. See USPTO, *Examination Guidelines for Determining Obviousness under 35 U.S.C. §103* (last modified Sep. 13, 2012), available at <http://www.uspto.gov/web/offices/pac/mpep/s2141.html>.
31. Christopher A. Cotropia & Mark A. Lemley, *Copying in Patent Law*, 87 N.C.L. REV. 1421, 1422, 1450–51 (2009) (proof of copying uncommon in patent infringement cases, except pharmaceuticals); Mark A. Lemley, *Should Patent Infringement Require Proof of Copying?*, 105 MICH. L. REV. 1525 (2007); Samson Vermont, *Independent Invention as a Defense to Patent Infringement*, 105 MICH. L.REV. 475 (2006).
32. 35 U.S.C. § 284. *In re Seagate Tech. LLC*, 497 F.3d 1360, 1368 (Fed. Cir. 2007) (en banc).

33. Cotropia and Lemley, *id.* at 1456.
34. *eBay, Inc. v. MercExchange*, 547 U.S. 388 (2006) (injunction as remedy for patent infringement not automatic). See *Ricoh Co., Ltd. v. Quanta Computer Inc.*, 2010 WL 1607908 (W.D.Wis. 2010) (denying injunction to non-practicing entity). However, the courts are not unanimous. See, e.g., *Callaway Golf Co. v. Acushnet Co.*, 585 F.Supp.2d 600 (D.Del. 2008) (granting permanent injunction even though patentee owner no longer manufactured the device incorporating the patent). See also *Paice LLC v. Toyota*, 504 F.3d 1193, 1314 (Fed. Cir. 2007) (noting possibility of ongoing royalty in lieu of injunction).
35. See Erik N. Hovenkamp, *Predatory Patent Litigation* (Northwestern Univ. Economics Working Paper, Aug. 5, 2013), available at SSRN, abstract #2308115.
36. See Hovenkamp, *Notice and Patent Remedies*, 88 TEX.L.REV. ONLINE 221 (2011), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1596789.
37. See, e.g., *Microsoft Corp. v. Motorola, Inc.*, 2013 WL 2111217 (W.D.Wash. Apr. 25, 2013). See Herbert Hovenkamp, “Competition in Information Technologies: Standards-Essential Patents, Non-Practicing Entities, and FRAND Bidding,” *INTERNATIONAL ANTITRUST LAW & POLICY* 439 (2013 Competition Law Institute, Fordham Univ. School of Law), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2154203.
38. *3M Innovative Properties Co. v. Tredegar Corp.*, 725 F.3d 1315, 1335 (Fed. Cir. 2013) (Plager, J., concurring).
39. *Id.* at 1336.
40. Cf. GUIDO CALABRESI, *THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* (1970) (when those about to have an accident cannot actually bargain to the most efficient result the legal system should try to predict the result of a free bargain, typically by assigning liability to the person able to avoid the accident at lowest cost).
41. See BOHANNAN & HOVENKAMP, *CREATION WITHOUT RESTRAINT*, *supra*, Ch. 33-59.

Legal Remedies For Patent Infringement: From General Principles To FRAND Obligations For Standard Essential Patents

BY RICHARD A. EPSTEIN¹ & DAVID J. KAPPOS²

At present, the traditional informal mechanisms for setting FRAND rates for SEPs have come under extensive attack by the Federal Trade Commission and elsewhere, from those who believe more limited damages and less frequent injunctions offer the best path to resolving disputes over Standard Essential Patents. In this article we take issue with those conclusions. We begin with an explanation of how a damage system for ordinary contract disputes does not typically rely on the three standard measures—expectation, reliance, restitution—for resolution, but uses liquidated damage to deal with the scenario where a defendant seeks to flout its own agreements. We argue that the techniques that work generally in contract law offer strong confirmation of the traditional rules of damages and injunctions now under wide assault.

I. SETTING THE STAGE: FROM COMMON LAW TO INTELLECTUAL PROPERTY

Patents live in two parallel universes. On the one hand, they are part of a larger system of property that includes as their closest relatives, land and chattels. On the other hand, patents exist solely as a creature of the state, which, through its general laws, allows individuals to obtain a patent—or the right to exclude others from making, using, offering for sale, and selling a patented invention for a limited period of years—so long as the patent application meets certain conditions the state sets as part of the patent bargain it offers to all prospective inventors. Property rights in land and chattels are typically created by initial occupation, which also confers an exclusive right to the first possessor, as well as subsequent purchasers. The role of the state is not to create the rights but to recognize and protect them. The bundle of property rights thus contains, in addition to the right to exclude, the right to possess, use and dispose of the property interest.³ That initial assignment of rights is usually not the last, for the original owner is then entitled either to use the property or to dispose of it to some third party, either by way of an outright assignment that transfers the property to another or by way of a limited transfer that leaves the original owner with a retained interest. The processes that allow for one such transfer allow for others as well. Any removal of a stick from the initial bundle of rights can impair the value of the owned asset for one of two reasons. Either it creates a system that leaves certain rights beyond exploitation, or it creates a built-in hold-out problem that requires the diminished property owner to enter into a transaction with either the state or some third party who for no good reason is put into a position

ANY REMOVAL OF A STICK FROM THE INITIAL BUNDLE OF RIGHTS CAN IMPAIR THE VALUE OF THE OWNED ASSET FOR ONE OF TWO REASONS. EITHER IT CREATES A SYSTEM THAT LEAVES CERTAIN RIGHTS BEYOND EXPLOITATION, OR IT CREATES A BUILT-IN HOLD-OUT PROBLEM THAT REQUIRES THE DIMINISHED PROPERTY OWNER TO ENTER INTO A TRANSACTION WITH EITHER THE STATE OR SOME THIRD PARTY WHO FOR NO GOOD REASON IS PUT INTO A POSITION TO BLOCK THE DEPLOYMENT OF AN INTEREST THAT IT CANNOT USE.

to block the deployment of an interest that it cannot use. The central task of creating an initial configuration of rights in any asset is to speed its path towards effective utilization with a minimum of impediments from outsiders is yet another application of the Coase theorem on the importance of low transaction costs as an engine of economic growth.⁴

Once these property rights are created by original acquisition, two other bodies of law come into play: contract and tort. First, the law of contract sets the ground rules by which property interests can either be alienated or shared. The basic logic of the contract system is to facilitate gains from trade, which can then be shared by both parties. Those transactions will take place so long as the joint gains exceed the combined transaction costs to the relevant parties. The legal system does not seek to identify for any private party the potential source of economic advantage, which varies by asset and trading partner. But what it does is establish standard modes of transacting that reduce the friction for voluntary transactions, in turn increasing the probability of their occurrence and the gains they generate.

In dealing with these contract rules it is necessary to deal with both a general case and a discrete problem. In the common competitive market system, the essence of a working market is one that allows each party the *absolute* right to refuse to deal with any one else. This rule in effect allows parties to fashion their own bargains, each knowing that if it pushes too hard for the ideal agreement it will lose its prospective trading partner to a competitor. But in those cases where the contracts in question involve a single supplier, such alternatives are not available, so the common law from the earliest time imposed a duty on the party that was the sole supplier of single services to provide them on what are now called fair, reasonable and nondiscriminatory, or FRAND terms.⁵ These monopoly situations have long been understood as exceptions to the usual rule of freedom of contract, and a complete analysis of any system of private relations requires special treatment to determine how these sole supplier arrangements operate in different contexts, including traditional common carriers, traditional public utilities and of course modern patent standard setting agreements.

The backstop to these basic contract principles is found in the law of tort, which is designed to prevent parties from circumventing the contract rules, whether in competitive or FRAND type situations, simply by taking those things they want without paying for them. In these situations, the conventional legal wisdom gives strong protections against the willful conversion of goods or the willful dispossession of property. In order to reduce incentives to bypass the contract system, the rules usually specify a return of the land or chattel, coupled with an award of interim damages. The simple impulse behind these rules is that people will avoid circumventing the established system of property rights if their unilateral actions leave them no better, and preferably worse off, than they were before.

Both contract and tort law present questions of optimal design, and it is on this point that private contracting practices enter. Sophisticated transacting parties choose rules that from start to finish will, from the *ex ante* perspective, maximize the joint gain of the parties. The task requires designing rules on liability, defenses and damages which in combination maximize the prospect of orderly performance, which those rules can do only by minimizing the probability and severity of breach. Understanding how contract provisions in competitive markets address issues of liability and damages offers a sensible template for developing the analogous rules needed to minimize the social dislocation that arises in stranger cases, most commonly by the taking or damaging of property. Once this pattern is understood, it is easier to devise the optimal rules for licensing patents and for addressing cases of infringement by strangers. For both tangible and intellectual property, the contractual

THE CURRENT SYSTEM, HOWEVER, UNTIDY, HAS WORKED WELL OVER MANY YEARS, AND THE NEW ROUND OF LITIGATION NOW WORKING ITS WAY THROUGH THE COURTS SHOWS THE PERILS ARISING WHEN COURTS EITHER TRY TO SET FRAND RATES OR RESOLVE DAMAGE ISSUES INVOLVING INFRINGEMENT OF SEP PATENTS.

rules offer an instructive lens through which to evaluate the tort rules. Accordingly, this article proceeds as follows. It first develops basic principles for analyzing contract liability and damages in the general case. It then considers torts, and determines how tort rules can be framed to work both as a complement and analogy to basic contract principles. With the two common law systems understood it is then possible thereafter to consider FRAND situations in the general case, which in turn sets the stage for a closer examination of possible solutions to FRAND issues, including various forms

of dispute resolution applicable to the modern patent system. In our view, the conclusion to be drawn is that the current system which relies heavily on informal mechanisms of dispute resolution that take into account the many cross-currents of the law of remedies and damages will work far better than any systematic effort to judicialize or otherwise formalize the dispute resolution process in connection with FRAND-encumbered patents. The current system, however, untidy, has worked well over many years, and the new round of litigation now working its way through the courts shows the perils arising when courts either try to set FRAND rates or resolve damage issues involving infringement of SEPs.

II. CONTRACTS: FROM FORMATION THROUGH BREACH

A. *Formation*

A full economic analysis of contractual design and structure must look at all stages of contract formation and performance. The actions taken at the time of contract formation will influence the behavior of the two (or more) parties to any given transaction. The rules governing the back end on the occasion of nonperformance will influence the behavior of both parties prior to and after the contractual relationship falls into distress. It is therefore important that one introduce the correct formalities, such as those required under the statute of frauds or recording statutes, so that parties know where they stand relative to each other and to the rest of the world.⁶ Within the field of patents, these formalities are substantial in connection with both claim construction and recordation. For both land and patents, the requisite formalities create an evidentiary record that reduces uncertainty over the life of the contract, helping to avoid or resolve potential disputes that may arise during the course of performance. Lower uncertainty reduces the probability of breach in the first place, by reducing the anticipated return to the breaching party. And when breach does occur, either by accident or on purpose, the initial formalities allow for more accurate dispute resolution at a lower administrative cost—a dual gain of enormous importance. In many cases, the transaction cost savings at the back end justify the higher costs incurred at the front end of the arrangement, thereby increasing the velocity of transactions, and with it overall social gain. A parallel argument can be made for recordation statutes that increase the security of transactions by allowing parties to know that they are indeed transacting with the owner of the property and not an imposter.

B. Breach

1. FULLER AND PERDUE

For these purposes, however, the key dimension of the dispute concerns the choice of back end remedies, assuming that the front end issues have been successfully resolved. Here again the analysis in contract and tort law has a heavy influence on how those issues are perceived in the law of patents. On this question there is a common perception, dating back to the major contribution of Lon Fuller and William Perdue in their article, “The Reliance Interest in Contract Damages”⁷ that damage options are set by the courts and not by the parties. But there is no reason why in a freedom-of-contract system the parties should not have broad discretion on both ends. The key issue for all parties is to determine the expected payoffs from either performance or breach. These payoffs depend on both the probability of the wrong, and the severity of the loss once it is committed. If the law limits the parties to discussing only the first but not the second, the parties lose flexibility in designing their own liability regimes. The dangers from limiting the parties’ flexibility are substantial because any initial discussion and concomitant actions tend to force the damages into certain preconceived categories. For Fuller these were limited to three major variations of one basic theme: restitution damages, by which the party in breach must return whatever cash or property he received from the innocent party; reliance damages, whereby the innocent party recovers all expenses incurred on the faith of the contract even if those expenditures provided no benefit to the defendant; and expectation of damages, by which the breaching party must put the innocent party in the same position, to the extent money can accomplish this, as he would have been in the absence of breach.

Central to the Fuller and Perdue position is that these various remedies form a hierarchy with restitution at the bottom, reliance in the middle, and expectation damages at the top. So stated, the scheme is subject to a number of powerful objections. First, the simple account of the different measures of damages offers no explanation of which remedy should be applied in which case. Nor is it possible to come up with such a theory on the Fuller and Perdue assumption because their analysis offers no coherent account of what parties seek to maximize by their choice of remedy. Instead they content themselves with Aristotelian references to the relative weight of the various interests.⁸ Absent that formulation the pairing of remedies with particular cases becomes at best an arbitrary procedure. To make matters worse, there is no a priori reason why the remedies that work to maximize joint cases should be confined to the three put on this narrow list for reasons that are more esthetic than functional. As will become clear later on, in dealing with licenses as well as other contracts, there are two disparate situations that require very different treatments. The first of these deals with the recurrent question of consequential damages that follow from the defendant’s nonperformance. The second deals with deliberate breaches of contract by defendants. In the first case, the task is to insure clarity of incentives for both parties in cases of sequential performance. In the second case, the central task is to make sure that strategic behavior by defendants does not subvert the system of voluntary exchanges. The former relies heavily on liquidated damages that *understate* the level of plaintiff’s loss; the second, on the use of injunctions and damages that may well *overstate* the measure of the plaintiff’s loss. The recent debate in the patent literature is over the second of these two issues, stemming from the seminal Supreme Court decision in *eBay v. MercExchange*.⁹ The importance of this distinction relies on a more complete analysis of both situations.

2. BARGAINING AFTER BREACH

One of the most common situations in the law of contracts involves a defendant’s breach of contract that leaves

open choices to the plaintiff on how best to respond to the loss. The social task in this case is to minimize the sum of the losses attributable to the breach, the steps taken to reduce that loss and the administrative costs of running the overall system.¹⁰ That system likely dominates in practice any of Fuller and Perdue's three preferred remedies. Getting the right answer does not necessarily speak to either high or low damages. It only addresses the relevant considerations for accomplishing two key tasks: first, designing, as a drafting matter, optimal remedies, and second, filling in the gaps on remedies when the parties are silent.

A fully informed calculation of damages requires addressing possible opportunism at all stages of the venture, and designing legal remedies to forestall them. In many real world situations, the parties do not think about the remedial phase at the outset, so the proper response is for courts to imply those additional terms, to the extent that they are able, in order to create correct incentives for cooperative behavior at low administrative cost. This process, moreover, does not translate into a prescription to award high damages in all cases. Much depends on context.

To understand when lower damages offer the best solution, it is instructive to look briefly at *Groves v. John Wunder Co.*,¹¹ where the court had to decide whether the correct measure of damages for breach of contract was the choice between the cost of completion of given work on the one hand, or the diminution of value of the subject property from the noncompletion of work on the other. There is no universal answer to this question, but in *Groves*, the lower measure is correct. In that case the defendant agreed to return land on which he had mined sand and gravel "at a uniform grade, substantially the same as the grade now existing at the roadway... on said premises." The cost of honoring that commitment was \$60,000, but the increase in market value of the land from honoring it was only \$12,000. There was no subjective value in raising the possibility that value in use is higher than value in exchange. Over spirited dissent the Court took the high moral ground and held that the larger sum was required because the strong nature of the basic contractual commitment meant that the defendant had to do what he had promised or face the consequences.

To appreciate why there is something amiss with this result, ask what happens if the defendant had no obligation to repair the land before he vacated possession. At that point, the owner would leave the property as is, instead of expending \$60,000 to secure a \$12,000 benefit. If that is how the owner would spend his own money, what makes it efficient for him to compel an expense that he would not make himself? A good set of remedies in the ex post state of the world should reach the quickest resolution of the underlying problem while avoiding the risk of strategic behavior by either side. Yet just that happens when the defendant is desirous of minimizing his financial burden, which he knows will be at least \$12,000. So to gain leverage, he does not meekly hand over \$60,000. Instead he announces that he is ready and able to perform the obligation for \$60,000, so as to tender perfect performance the plaintiff does not want. Now the defendant's threat sets up a bargaining game in which both sides are better off if they pick some cash transfer payment above \$12,000 and below \$60,000 to liquidate the performance obligation.

A moment's reflection, moreover, yields two further difficulties. First, there is no unique point within that range, satisfying the requirement of joint improvement. All figures between \$59,999 and \$12,001 satisfy the condition. Second, efforts by both sides to capture that surplus will necessarily consume resources, shrinking the surplus the two parties have to share. In a noncooperative game rife with bluffing, that loss could be quite great. This entire bargaining scenario is a pure negative sum game, for there is no allocative gain, since neither side wants to return the land to its original contour. It takes little imagination to conclude that in the ex ante

state of the world neither side would adopt a measure of damages that invites this ex post bargaining game. Using the decline of market value as the measure of damages eliminates that risk in its entirety, leaving only an unavoidable valuation question that has to be answered under both measures of damages. It is also worth noting that if the decline in market value is greater than the cost of completion, the damage rule will never come into play because defendant will find it easier to just complete the work, so that the diminution in value rule is robust in all states of the world, in a way in which the cost of completion rule is not. There are, then, cases where it seems clear the lower remedy is preferable from the ex ante perspective. There are also cases, such as those involving the design of a new home, where the subjective value to a property owner can easily exceed the cost of completion, at which point the cost of completion becomes the preferred measure of damages.¹²

THIS IS ONE OF THE MAJOR QUESTIONS IN PATENT DISPUTES, ONCE IT IS DECIDED THAT INJUNCTIVE REMEDIES WILL NOT ISSUE. IN EFFECT, THE RULE ON EFFICIENT BREACH GIVES THE INDIVIDUAL USER OF A PATENT AN OPTION AT AN UNLIQUIDATED PRICE TO DECIDE WHETHER TO TAKE A LICENSE FROM THE PATENTEE OR SIMPLY INFRINGE, KNOWING THAT A DAMAGES ACTION MAY BE BROUGHT AGAINST IT.

In other situations, it is possible to avoid any evaluation of consequential damages by following the U.C.C. rule that disallows consequential damages to any plaintiff who is in a position to cover with identical goods after breach.¹³ Why allow for those damages when there is a perfect mitigation strategy that sidesteps the complex cost benefit analysis involved in all too many mitigation cases?

3. EFFICIENT, OR NOT-SO-EFFICIENT, BREACH

The second issue in these damages cases is whether to use an expectation measure of damages in the face of a deliberate breach by a defendant who thinks that his

profit is greater than his exposure to damages. This is one of the major questions in patent disputes, once it is decided that injunctive remedies will not issue. In effect, the rule on efficient breach gives the individual user of a patent an option at an unliquidated price to decide whether to take a license from the patentee or simply infringe, knowing that a damages action may be brought against it. In some cases, those risks will be deemed too high, so that the license will be sought. But in other cases (often dealing with different users of the same patented technology) the results will vary by user of the technology. Some of these parties will find themselves constrained to choose the license, while others may move sharply in the opposite direction. It is impossible in the abstract to trace the motivations of the various players, which depend on their exposure to suit, their asset base, the background availability of other technologies should the cost prove too high, and many other variables. But nothing is in practice more common than to see different paths pursued by different parties.

It is in these cases that the weaknesses of Fuller and Perdue's reliance and expectation damages become clear. One popular approach that dates back to work done by Charles Goetz and Robert Scott¹⁴ takes the view that the hidden virtue of the expectation measure of damages is that it induces individuals to breach their promises only in those situations where the alternative use of the resource is of greater value than the promised use. That view rests on the assumption that there is an overall Pareto improvement so long as the defendant is able to make the plaintiff whole by paying the requisite damages, even if he keeps all the gain to himself. But it ignores the possibility, in both ordinary contract and patent licensing cases, that the lost profits—the

consequential damages from breach—may not be known to either party, at which point the approach loses its utility, both for its substantive uncertainty and high administrative expenses. The situation only gets worse, because any strategy that looks solely to the incentives of the defendant to breach, ignores the risk that both parties can engage in opportunistic behavior once a transaction goes off the rails.

The first rule for controlling bilateral forms of misbehavior is to make sure the remedies afforded to each party are independent of the conduct of the other party, so that both sides face correct incentives on the key choice of whether to breach or perform. To see why this is necessary, consider a case where the defendant under a hotel construction contract is required to complete building by a certain date after which the plaintiff is entitled to use the hotel as it thinks best.¹⁵ Neither the reliance nor expectation damage formulas make sense. If the defendant is on the hook to pay reliance costs, there is a strong incentive on the plaintiff to make extra expenditures knowing that the defendant must act as a guarantor of the loss if completion is late. But switch to the expectation measure of damages, and the plaintiff has a strong incentive to enter into multiple transactions that offer promise of great gains, knowing that these gains too are guaranteed in the event the facility is not finished in time. Both measures of damages allow the plaintiff to speculate at the expense of the defendant and to act in ways it would not if it had operated as a single owner in the construction and operation of the hotel.¹⁶

STATED IN THIS FORM, THE THEORY HAS TO BE INCORRECT. THERE IS NO WAY IT IS EFFICIENT TO IMPOSE BY OPERATION OF LAW A STANDARD OF DAMAGES THAT IS NEVER ADOPTED IN ANY VOLUNTARY TRANSACTION, BE IT IN PATENTS OR ANYWHERE ELSE. INDEED THE PRECISE OPPOSITE IS TRUE. THE LAW IS FILLED WITH ALL SORTS OF BUSINESS ARRANGEMENTS IN WHICH OPTIONS ARE EXPRESSLY CONFERRED.

The challenge here is to create a set of damage rules that replicate the incentives facing the single owner, for whom complications of externalities and holdouts are eliminated by virtue of his control over the relevant inputs and outputs from any decision he takes. The issue—and this is critical—for the single owner situation is the need to devise a strategy that

will incentivize him to maximize subjective utility. The calculations are often difficult, but by the same token the process is not infected by incentive or bargaining problems.

The question thus arises: how can the single-owner approach help structure cooperative contracting ventures when the theory of efficient breach leads systematically to the wrong results? On this issue it is instructive to start by recalling Oliver Wendell Holmes's famous aphorism, that "the only universal consequence of a legally binding promise is, that the law makes the promisor pay damages if the promised event does not come to pass."¹⁷ Stated in terms of the modern theory of efficient breach, this statement morphs into the proposition that at common law a promise leaves the promisor with the *option* to perform or pay damages. Stated in this form, the theory has to be incorrect. There is no way it is efficient to impose by operation of law a standard of damages that is *never* adopted in any voluntary transaction, be it in patents or anywhere else. Indeed the precise opposite is true. The law is filled with all sorts of business arrangements in which options are expressly conferred. In *none* of those cases is the option price expressed in terms of an unliquidated amount that requires an extensive dispute resolution process, either public or private, long after the dispute is over, to divine the correct option price. Instead across multiple settings, any contractually chosen option takes the form of a specific number, or

of a formula that converts easily into a number, with the addition of a few public facts such as the age of the party or the number of days or hours or seconds a party is in breach.

Given this constraint, legal theory confirms why such formulas should be respected by the law. They are easy to adjudicate after the fact and thus add real value to the parties by eliminating one dimension of uncertainty in the ex post state of the world. It is for this reason that specific (in the sense just mentioned) numbers are preferred to formulas that seek to place the plaintiff in the position she would have enjoyed had the defendant fulfilled its promise, which gives no precise answer. The choice of the dollar figure used in the event of a break-up is not something the legal system can supply in general, any more than it can determine in the abstract the price of all goods and services held up for sale. What it does is identify the ideal type of measure and then leaves it to the parties to fill in the blanks by adding the appropriate number or numbers.

4. LIQUIDATED DAMAGES

At this point the focus of the legal system should be on liquidated damages. In practice, whether private parties take up the invitation to liquidated damages is a separate question that often depends on the size of the transaction. The larger the transaction, the more likely they will come to an agreement on those numbers, as is commonly the case with executive severance packages or break-up fees in connection in large and complex corporate mergers and acquisitions. The only time one resorts to a “reasonable price”¹⁸ is for a completed transaction when specific goods sold are consumed or resold, when no price term has been set in advance. Better that determination than the failure to have any remedy at all as a result of calling a contract “indefinite” because it does not contain a suitable pricing term.¹⁹

In light of the above, the strong presumption should be in favor of the liquidated damage provision even if it does not meet the standard Restatement and Uniform Commercial Code requirement to offer some “genuine pre-estimate” of the actual losses suffered.²⁰ It is simply not appropriate that the sole, or even major function of the liquidated damage clause is to estimate future damages where they are difficult to calculate.²¹

It is also common in major commercial transactions to use liquidated damages to structure the rules governing breach to reduce both the probability of a contractual breach and the dislocations it causes. In dealing with this issue the expectation measure is routinely displaced by explicit provisions offering far better incentives that inform how sequential performance, such as the hotel construction contract, should be analyzed. In the simplest of situations, the construction company performs first and the hotel owner makes its decision on how best to deploy the property when the property is under construction. An efficient system has to accomplish two tasks simultaneously. The first is to make sure the defendant does not profit from its own wrong, so as to reduce the likelihood of tardy performance. The second is to make sure the plaintiff does not speculate at the expense of the defendant by taking measures in mitigation that it would not take if it were a single firm having both construction and operations divisions.

The way to accomplish both tasks is through a *liquidated damage* provision that indicates the amount the defendant has to pay in the event of lateness as a function of time, as with the construction contract discussed above. This task requires at a minimum that the parties set a price schedule, which could be linear, or which could increase or decrease with time in accordance with some set plan. The first element of gain from this approach is that it gives the construction company incentives to perform so long as the penalty in question leaves it

worse off than with nonperformance, which is typically the case so long as the obligation to complete construction remains even after the contract is late. The second advantage is that it removes temptation for a plaintiff to mitigate losses inefficiently, knowing that the defendant will have to pay the bill. The fixed figure eliminates conflict of interest because the fixed sum paid by the defendant has no influence on any future decision made by the plaintiff, as the defendant pays the same amount no matter what the plaintiff does. The plaintiff thus has no incentive to spend either too little or too much on future remediation, or ongoing commitments with third parties. One simple instrument thus handles correctly these sequential performance cases far better than any formula that first imposes on a plaintiff the duty to mitigate, which it then hedges with limitations that require the plaintiff's steps to be "reasonable" in light of his endowments and circumstances.

The situation gets more complicated when one relaxes the initial assumption of sequential performance and has interlocking obligations for performance and payment, or multiple performance obligations sequenced between the parties, which often happens in installment transactions. No formula can cover all these cases, so additional governance mechanisms are needed to secure performance by all parties. Often times some independent party determines full compensation on breach, without allowing the innocent party to exercise its holdout potential. As will become clear shortly, this general point applies to FRAND (fair, reasonable and nondiscriminatory rates) to mitigate the holdout problem. There is always a level of litigation potential in these settings, but the informal sanctions that constrain the process—the repeat dealings that extend from present to future projects; the use of architects and others as independent mediators, the development of a body of industry practices that cover the most recurrent situations—reduce the frequency and intensity of disputes to reasonable levels.

III. DAMAGES AND INJUNCTIONS IN PATENT LICENSING CASES

A. *Incremental Damages under the 2011 FTC Report-The Role of FRAND Standards*

The inescapable conclusion from the arguments in the previous section is that the theory of efficient breach is inefficient by its own standards. In voluntary markets no one uses that approach—not in intellectual property markets such as standard-setting, and not in other contexts. This point also calls into question the recommendation of the Federal Trade Commission with respect to damage remedies for standard essential patents, and provides insight into the difficulties associated with the litigation over FRAND standards now making its way through the courts. Let us take up these two points in order.

The FTC Recommendation report contains this key passage:

“Recommendation. Courts should apply the hypothetical negotiation framework to determine reasonable royalty damages for a patent subject to a RAND commitment. Courts should cap the royalty at the incremental value of the patented technology over alternatives available at the time the standard was defined.”

The difficulties with this proposed rubric are legion.²² In the first instance, FRAND patents are difficult to evaluate with any of the standard techniques used for stand-alone patents. The value of any given patent included in a standard depends on its interaction with other patents included in the standard. The technical committees that work on these issues have to consider a wide range of possible permutations, so it is unlikely

THE INESCAPABLE CONCLUSION FROM THE ARGUMENTS IN THE PREVIOUS SECTION IS THAT THE THEORY OF EFFICIENT BREACH IS INEFFICIENT BY ITS OWN STANDARDS. IN VOLUNTARY MARKETS NO ONE USES THAT APPROACH—NOT IN INTELLECTUAL PROPERTY MARKETS SUCH AS STANDARD-SETTING, AND NOT IN OTHER CONTEXTS. THIS POINT ALSO CALLS INTO QUESTION THE RECOMMENDATION OF THE FEDERAL TRADE COMMISSION WITH RESPECT TO DAMAGE REMEDIES FOR STANDARD ESSENTIAL PATENTS...

the elimination of one particular patent from the standard can be cured by substituting some other patent with identical functionality in its place. Any effort to reconfigure these SEPs months or years after they are first put into place is fraught with both conceptual and practical difficulties, which do not get any easier when licensees unilaterally deviate from FRAND terms without consent, which in practice turns out to be the most likely possibility.

The FTC assumes that the only function of patent damages is to make sure that the patent holder gets at most the amount of money it could have gotten relative to a patent's next best alternative. In making this particular calculation, however, the FTC ignores a number of critical points. Voluntary compliance with a patented standard comes at little or no cost to the firm. But the moment potential standard users know

their damages are capped at “the incremental value of the patented technology over alternatives available at the time the standard was defined,” they have an incentive to opt out from the voluntary market—that is to engage in *deliberate* infringement of the applicable patent in order to improve their own position. Setting the damage cap at actual damages reduces the incentive, and therefore the likelihood of taking licenses. Why comply, if infringement will on average leave a particular party better off than joining into the standard? To avoid that risk, a patentee might reduce royalty rates below their optimal level. But that strategy comes at a cost because the reduced rate of return from licensing reduces returns on innovation, which in turn reduces the incentive to innovate. Yet this strategy may well make sense if the only alternative requires expensive and uncertain litigation. On this score, it is important to stress that no voluntary option would ever be tied to an unliquidated standard of damages. The pattern of defection if practiced by one can be replicated by others, at which point the internal governance structure needed to keep SSOs in place is subject to serious stress, for individual patent holders now face high costs of litigation, uncertainty of outcome, and delayed revenue streams on their patents, all of which lower the rate of return below what it would be in voluntary negotiations. Why would we think a standard of damages that no one adopts voluntarily, offers a solution to the problem of contracting over damages?

The second point is that the proposed rubric, which ties patent compensation to the incremental value at the time the standard was defined, creates a heads-I-win-tails-you-lose dynamic. If the value of the standard depreciates over time, the price will fall and the latecomer can reap the rewards of delaying. But by the same token if the standard increases in value, the holdout has in effect an option to sign on at the original price, given his credible threat to go without the license if the patent holder does not acquiesce in the original (lower) price. That free option to the putative licensee thus reduces the return to parties who set the standard as well as early good-faith adopters who pay for licensees, so the reward goes to parties who game the system, and not to those who contribute to its overall long term value.

The point here is not to pretend that the risk of holdups does not arise with respect to SSOs. The basic

invocation of the FRAND standard is only intelligible against the backdrop of a world in which the holdup problem is acknowledged. But the correct question involves the *relative imperfections* of the alternative institutional arrangements, and on that question there is no systematic engagement by the FTC or, as far as we can tell by other defenders of the position taken in the FTC report. But there is no good reason to think the holdup problems associated with FRAND patents require extensive litigation to resolve them. Holdup problems are common where businesses are required to modify an existing contract to take into account changed conditions, and in these situations, successful negotiations follow the pattern whereby the parties seek to cost-justify their demands for increased prices by identifying the changes in cost structure that call for alteration. The effort to make accurate cost determinations is the proper way to negate the holdup potential that can arise in these situations. When parties follow that approach no economic duress occurs. But if the parties simply announce that they need compensation for the additional costs imposed on them by the other party's breach of contract, the holdup will constitute economic duress if there is no effort to cost-justify the increases.²³

Historically, it is just this form of analysis that governs the regulation of monopoly power for public utilities in connection with power and telecommunications services that throughout their history have been subject to FRAND regulation. In each of these cases, the entity in question ran major facilities over which some branch of government had the power to set rates that were intended to walk the fine line between the toleration of monopoly profits on the one hand and the confiscation of capital on the other. The size of the rate base in these cases was huge; often the pace of technological progress was relatively slow; the cost of administration through public utility commissions was small compared to the revenues that passed through the system. The formal administrative processes invoked in these cases, subject to both administrative and constitutional oversight, may well have been less than ideal, but the basic pattern withstood these challenges for close to a century after the issue first raised its head with the passage of the Interstate Commerce Act of 1887.

THE POINT HERE IS NOT TO PRETEND THAT THE RISK OF HOLDUPS DOES NOT ARISE WITH RESPECT TO SSOs. THE BASIC INVOCATION OF THE FRAND STANDARD IS ONLY INTELLIGIBLE AGAINST THE BACKDROP OF A WORLD IN WHICH THE HOLDUP PROBLEM IS ACKNOWLEDGED. BUT THE CORRECT QUESTION INVOLVES THE *RELATIVE IMPERFECTIONS* OF THE ALTERNATIVE INSTITUTIONAL ARRANGEMENTS

The situation with respect to FRAND patents in many commercial sectors in current times is entirely different. The technology moves quickly, so standards have to be frequently revised or abandoned in light of technological change. Quick and easy resolution of cases is critical. At the same time, the repeat-play nature of these games tends to place clear limitations on how holders of SEPs behave relative to other institutions. The familiarity from repeat play improves the odds that those companies committing their patents to SSOs will be less likely to strategically opt out of the system. Indeed, in the patent context, it is rare that a single standard will govern in all cases.²⁴ Thus it may well be that many different patents and standards are in play at the same time. The licensees in some cases may be in a position to supply cross-licenses that could easily justify a downward adjustment in pricing policy. Or some licensees may provide a useful service to the SSO, which again warrants a price adjustment. Nor is there any reason why all licenses must involve simple royalty arrangements: fixed payments might well be part of the dynamic.

An item to note is that currently there are a large number of SSOs that crank out standards for new products and update standards for old ones. If the coordination problem were as serious as suggested, we should see routine breakdowns in SSOs, yet we do not see that at all in practice. It is worth noting that the FTC attached no weight to the testimony of all major SSOs that their activities are far more uneventful than popular critique

THIS POINT IS CRITICAL BECAUSE IT IS NOT POSSIBLE IN THE ABSTRACT TO DECIDE WHICH PARTY IS WEARING THE WHITE OR THE BLACK HAT. HOLD-UPS ARE NOT A ONE-SIDED PHENOMENON, BUT CAN BE PRACTICED BY AN ASTUTE LICENSEE THAT BARGAINS AGGRESSIVELY JUST AS THEY CAN BE PRACTICED BY A LICENSOR.

conjures. These organizations use a loose version of the FRAND obligation to guide their negotiations, but they do not lock themselves into any strict formula, given the variety of situations they face.

This point is critical because it is not possible in the abstract to decide which party is wearing the white or the black hat. Hold-ups are not a one-sided phenomenon, but can be practiced by an astute licensee that bargains aggressively just as they can be practiced by a licensor. One point to note, however, is that any regime treating litigation as a first-best alternative has

to beware of the high costs and the enormous delays of that practice.

B. Recent Litigation Under the FRAND Standard

This analysis is borne out by an examination of two recent FRAND cases:

1. MICROSOFT CORPORATION V. MOTOROLA

The first illustration of potential problems is *Microsoft Corporation v. Motorola, Inc.*,²⁵ before Judge James J. Robart, which involved a dispute between two titans over standards, when Microsoft and Motorola were unable to agree to terms whereby Microsoft could obtain licenses to Motorola SEPs. Both sides agreed that FRAND rules applied and both noted that the purposes of the rules and procedures were to enable technical experts to devise the best standard with the lowest possible legal drag. But notwithstanding that strong agreement with respect to basic points, the two sides brought into the fray a combined 18 expert witnesses to opine about all aspects of the case before Judge Robart, who found each and every one of them credible. During the course of his decision, Judge Robart started the analysis with the oft-cited decision of *Georgia-Pacific v United States Plywood Corp.*,²⁶ which lists fifteen separate factors that are relevant, but not dispositive on the royalty question. It is accepted that courts have “wide discretion” on how Georgia Pacific factors are applied.²⁷

The concern here is that cases like this are luxuries that no legal system can easily afford given the huge number of standards that must be set on a near daily basis. The wide range of factors suggests there could be a wide range of acceptable outcomes. We see no obvious way the rules of adjudication can be made crisper and clearer, which counsels toward a regime favoring the use of informal processes whose outcomes are accepted even if they may be less than ideal. There is simply no reason to believe that a judicial outcome in a formal setting will do any better, let alone sufficiently better to justify the enormous cost and delay. Litigation simply cannot avoid using the *Georgia Pacific* factors, as clumsy as they are. It is for just this reason that every effort should be made to strengthen industry practices that yield agreement in high fraction of standard-setting disputes.

2. IN RE INNOVATIO LP

i. Background

That conclusion is not altered by a review of the second recent SSO case *In Re Innovatio LP*,²⁸ which also involves a FRAND dispute, in the context of a suit for patent infringement, also relating to the WiFi standards promulgated in 1997. In his decision, Judge Holderman of the Northern District of Illinois separated out the FRAND issues from the other issues in the case, and once again revealed the immense difficulties that come from adjudicating SEP disputes in Court. In this instance, Innovatio took assignments from various parties, all of whom had made FRAND commitments during the Institute of electrical and Electronics Engineers (IEEE) standard setting process. The Court's decision dealt only with damages, bracketing for the moment any issue of infringement. By virtue of this new posture, Judge Holderman concluded that he had to set a precise estimate of damages, not just the "reasonable royalty range" appropriate to the determination in *Microsoft v. Motorola*, which involved not infringement, but broken-down negotiations.

LITIGATION SIMPLY CANNOT AVOID USING THE *GEORGIA PACIFIC* FACTORS, AS CLUMSY AS THEY ARE. IT IS FOR JUST THIS REASON THAT EVERY EFFORT SHOULD BE MADE TO STRENGTHEN INDUSTRY PRACTICES THAT YIELD AGREEMENT IN HIGH FRACTION OF STANDARD-SETTING DISPUTES.

In setting damages, Judge Holderman examined testimony from a large squad of 10 experts, five on each side, to whom he did not attach equal credibility. As with *Microsoft v. Motorola*, the merits of the standards dispute are not the focus here. What matters are the assumptions made about how the case should be decided within the FRAND framework. On that question three issues require attention: patent validity, reverse hold-up, and product base with patent stacking.

ii. Patent Validity

One issue in these cases is how we should take into account doubts about the validity of the patent in determining FRAND rates. The correct benchmark is how that issue is dealt with in voluntary negotiations. But the *Innovatio* decision gives no inkling as to what that practice might be. It would therefore be helpful to know whether the SSOs screen first for patent validity or whether they simply assume validity based on the examination conducted by the USPTO determination that carries with it some positive error rate. In dealing with this issue the court notes that its calculation of FRAND rates assumes perfect validity of the patent.

The above validity assumption leads to an important objection to the adjudication process. Why should an uncertainty be ignored when its resolution is critical to simulating a "hypothetical bilateral negotiation" between the parties?²⁹ One common criticism of tort damages and patent damages is that they do not reduce liability to take into account the uncertainty that is eliminated where the plaintiff prevails by a simple preponderance of evidence.³⁰ That objection is not decisive because of the equal and opposite bias whereby a plaintiff recovers nothing in a case where the probability of defendant's wrong is positive, but less than 50 percent. This cancellation approach assumes the probability distribution is symmetric around a mean of 50 percent, which need not prove true in ordinary two party disputes. But in dealing with disputes involving multiple plaintiffs

in identical positions, the preponderance of the evidence standard runs the risk of significant underdeterrence. If liability in each of 100 identical cases can be established at 40 percent, the correct level of deterrence is not obtained by marking them all down to zero.³¹ Nor on the opposite end is optimal deterrence obtained if each of 100 cases in which liability is established at a 60 percent probability is awarded full damages. Ironically, settlements that pick an intermediate position are not always able to soften the errors of over and under deterrence. Any such power of that self-correction device depends on the liability rule established. Thus if it is certain that liability carries 60 percent probability, settlement will be at 100 percent, given that the parties bargain in the shadow of the law.

For the *Innovatio* case, the assumption of perfect validity leads invariably to under valuation, as more patents than should be are included in the damages base, leading to a lower recovery per patent. With no evident way to correct for this in the validity proceedings to follow, the patents ultimately found valid will receive less than their share of recoveries.

iii. Reverse Hold-Up

A second point of difficulty with the *Innovatio* decision relates to what Judge Holderman called the problem of the “reverse hold-up,” which is the concern that implementers of a standard will infringe standard-essential patents without taking a license, thereby forcing innovators to engage in costly litigation before realizing the value of their inventions.³² Judge Holderman was “not persuaded that reverse hold-up is a significant concern in general, as it is not unique to standard-essential patents.”

The court may be correct that historically and up to current-day, reverse hold-up has not been a significant problem, at least not to the extent of entering the jurisprudence. But one reason for this salutary situation is that historically and up to current-day informal negotiations have worked well in the shadow of injunctive relief.³³ Nonetheless, the court’s logic can be faulted on two grounds in this instance. The first is that merely because the reverse hold-up problem is generalized beyond SSOs does not mean it is insignificant within the field. Quite the opposite: its general importance reinforces the concern in this particular area where the need for coordination across many parties makes the reverse hold-up problem more and not less important. Second, the frequency of the reverse hold-up problem is not exogenous to the legal system, but heavily depends on the legal rules that govern these negotiations. The historical data arises both before and after the *eBay* decision. In the former period, issuing an injunction as a matter of course subject to narrow defenses based on such matters as laches and estoppel, is calculated to slow down this problem. The post-*eBay* rules are more lax.³⁴ The past therefore is not prologue to the future. We should be confident that any major change of remedies in the direction of the FTC recommendations will increase the risk of reverse hold-up. Given the cost and expense of litigation around SEPs, it should be clear that routine litigation, or even routine arbitration is not the answer.

FOR THE INNOVATIO CASE, THE ASSUMPTION OF PERFECT VALIDITY LEADS INVARIABLY TO UNDER VALUATION, AS MORE PATENTS THAN SHOULD BE ARE INCLUDED IN THE DAMAGES BASE, LEADING TO A LOWER RECOVERY PER PATENT. WITH NO EVIDENT WAY TO CORRECT FOR THIS IN THE VALIDITY PROCEEDINGS TO FOLLOW, THE PATENTS ULTIMATELY FOUND VALID WILL RECEIVE LESS THAN THEIR SHARE OF RECOVERIES.

iv. Patent Base and Royalty Stacking

The third issue of note in *In re Innovatio* is that setting the right royalty requires more than selection of a single number for the royalty. It also requires choosing the proper unit over which the royalty determination should be made. In dealing with this issue, Judge Holderman chose the “smallest salable patent practicing unit” and not some larger product of which that unit is a part.³⁵ The ostensible reason for this decision is the need to guard against the risk of overcompensation with respect to contributions that others make to the combined product, which could happen when one patent is “stacked” on top of another. The problem with this argument against patent stacking is that it focuses exclusively on one type of error and ignores a second. There is no a priori reason why the value of an SEP should be measured only by the smallest unit in which it is housed. In voluntary markets, the gains of a particular device will be dependent at least in part on the extent it can be resold to downstream users. If parties attach different values to the device, price discrimination is a perfectly respectable form of behavior by which to capture those downstream rents so long as there is no horizontal collusion. To the extent these practices are observed with ordinary patents, they should be permissible with SEPs as well. It is an open question as to the extent SSOs should engage in price discussions in routine cases. But it is surely a mistake to rule out this kind of evidence on an a priori basis. So long as complications are present, it is a serious oversimplification of the problem to propose, as do Mark Lemley and Carl Shapiro, that disputes over SEPs can be settled by an arbitration mechanism, similar to that used in baseball, that looks only at the basic royalty arrangement.³⁶ That might work in some subset of cases, at which point the parties can adopt it voluntarily. But hasty generalization carries with it real dangers. Rate base discussions are highly important, and must always consider the twin perils of over and under compensation that is at stake not only in the public utility context, but also with SEPs.

THE HISTORICAL DATA ARISES BOTH BEFORE AND AFTER THE EBAY DECISION. IN THE FORMER PERIOD, ISSUING AN INJUNCTION AS A MATTER OF COURSE SUBJECT TO NARROW DEFENSES BASED ON SUCH MATTERS AS LACHES AND ESTOPPEL, IS CALCULATED TO SLOW DOWN THIS PROBLEM. THE POST-EBAY RULES ARE MORE LAX.³⁴ THE PAST THEREFORE IS NOT PROLOGUE TO THE FUTURE. WE SHOULD BE CONFIDENT THAT ANY MAJOR CHANGE OF REMEDIES IN THE DIRECTION OF THE FTC RECOMMENDATIONS WILL INCREASE THE RISK OF REVERSE HOLD-UP.

3. CASCADE EFFECTS

It is in our view also dangerous to think about the resolution of FRAND disputes in isolation from the larger system of patent licensing. That system must address not only the static two party disputes like *Microsoft v. Motorola*, but it must also take into account the point that FRAND negotiations face the same cascade problem that is familiar in other contractual settings. One risk of looking at contractual arrangements through litigated cases is that it can appear that all contracts are stand-alone arrangements between two parties who act in total isolation from the rest of the world. From the ex ante perspective nothing could be further from the truth. Whether one thinks of complex production line arrangements, or complex construction projects, a single general contractor is charged with making sure that each of the subcontractors performs on time and as required. Where that is done, each contract can build on those that came before and set the stage for those that come

afterwards. But where there is a break in the chain of performance, the effects ripple up and down the line of production. Downstream producers will claim that they are entitled to damages to off-set the costs needed to adjust to the prior breaches. The head contractor may well insist that the downstream producers took the risk of those delays or errors in taking their piece of the work; or the head contractor may seek to pass the costs back up the line to the party in breach. In intellectual property cases, one license can set the stage for a dozen

IT IS CRITICAL TO UNDERSTAND THAT THE USE OF ANY DAMAGE REMEDY CAN LEAD TO AN ENTIRE NETWORK OF CONTRACTS, THE ENTIRE IP SUPPLY CHAIN, UNRAVELING. THE NONPERFORMANCE OF ONE CONTRACT SETS THE STAGE FOR THE NONPERFORMANCE OF THE NEXT, AND SO ON DOWN THE LINE.

sub licenses, so the same lattice found in traditional businesses appears in this context.

No matter what the particular context, the pattern of analysis is the same, and more complex than dealing with two parties. It is critical to understand that the use of any damage remedy can lead to an entire network of contracts, the entire IP supply chain, unraveling. The nonperformance of one contract sets the stage for the nonperformance of the next, and so on down the line. It is always a disputed question whether the non-

performance amounts to a breach or is excused by the mistakes that occurred higher up the line. No sensible businessperson wants to incur the risk of sorting out these multiple claims on the fly, especially when work has to go on even if the individual disputes are not resolved. Whenever interlocking arrangements are involved, the gains of performance over breach increase exponentially. It is therefore in precisely these high-velocity transactions that the damage rules should be tailored to encourage performance, which cannot be done when they are calibrated to incremental breach.

One key mistake of the FTC report is that it ignores all these distinctive features of the patent market in choosing a standard that is nowhere found in current business practices. And it does so even though it cannot show any systematic failure in the current system. The Epstein-Kieff-Spulber article referred to the position of Qualcomm with respect to its own uneventful involvement with some 84 standard setting organizations.³⁷ That article noted that all of the major standards organizations reported no difficulties in going about their routine work, a view that would not be explainable if the coordination problems were as profound as the FTC suggested. We see no evidence that the basic analysis has changed over the past two or three years.

IV. THE TORT SETTING

Thus far we have examined how the law should analyze the licensing agreements patentees enter into with their various licensees. But in many of these situations, it is apparent that potential licensees may make the decision not to enter into any contract at all, and choose to simply infringe. Nothing is more common than for negotiations to break down for the use of someone else's property, be it land or patented technology. At that point, a party may decide to take its chances in carrying out its plans, knowing that down the road it may well face litigation that could result in potential exposure to liability.

The question then becomes how to structure legal remedies in response to that decision. The fulcrum of analysis is that information gleaned from voluntary transactions should be carried over, whenever possible. That result can only transpire, however, if the legal system can impose strong incentives by way of injunction or damages so that it is not likely for any potential trespasser or infringer to be better off skirting the legal system

rather than working through voluntary transactions. At this point it follows that if the damages are limited to the same measure of “incremental damages” that are the calling card of the FTC report, the incentives to infringe increase, precisely because of the same time delay, administrative costs and uncertain recoveries that occur when other kinds of licensees are able to breach their arrangements with impunity. In these cases, the object of stringent remedies is to forestall the occurrence of breach in the first place by making sure potential infringers come to the table to bargain within the framework of the FRAND obligations noted above.

These cases present very different issues from cases where the damage remedies are kept deliberately low in order to induce a plaintiff to mitigate losses after the defendant has breached its obligations. In those cases, of course, there is literally no reason to think about enjoining any behavior by either party because no one has sought to circumvent the outcome of voluntary transactions, which indeed are the source of the standard limitation on consequential damages. It is for just this reason that damages in product liability cases can be too large, because they downplay the role of plaintiff’s misuse in determining liability. The upshot is a vast expansion in liability with little or any improvement in safety, given the higher level of misconduct by product users.³⁸ But in cases where a stranger makes a conscious decision to use, without consent, the property of another, then a strong set of remedies is as relevant for patented technologies as it is for land or other tangible assets.

In the above settings, it is best to use clear rules to determine whether liability should be imposed, as it is generally unwise to search for some point on a continuous distribution, which acts as an on-off switch for personal liability.³⁹ A continuous set of points maps well into determination of damages as matters of degree that depend on the relative seriousness of the injury in question. The demarcation of these two functions works as well in patent infringement cases between strangers, where the question of liability is solved before the question of remedy is addressed.

AT THIS POINT IT FOLLOWS THAT IF THE DAMAGES ARE LIMITED TO THE SAME MEASURE OF “INCREMENTAL DAMAGES” THAT ARE THE CALLING CARD OF THE FTC REPORT, THE INCENTIVES TO INFRINGE INCREASE, PRECISELY BECAUSE OF THE SAME TIME DELAY, ADMINISTRATIVE COSTS AND UNCERTAIN RECOVERIES THAT OCCUR WHEN OTHER KINDS OF LICENSEES ARE ABLE TO BREACH THEIR ARRANGEMENTS WITH IMPUNITY.

Within this framework injunctive relief usually has no relevance to single, one-off situations that involve accidental harms alone. But it is a far different situation in the law of nuisance, for example, where adjacent property owners have a high probability of repeat interactions, all of which are deliberate in the sense that the defendant knows from prior experience that his conduct will necessarily produce, the next time, the harm it did on a previous occasion. In these settings, the standard remedy is injunctive relief, supplemented by a damage action that picks up two kinds of losses. First, the prior harm that was completed before the injunction has issued. Second, the commendable reluctance of courts to eliminate all background risks associated with certain activities on the ground where at the margin, the dislocation sustained by a defendant (who could be shut down if a single particle of coal dust escaped its operations) is far disproportionate to the trivial relief afforded the defendant. In light of this reluctance, injunctive relief can be introduced in stages, and coupled with clean-up damages. In the nuisance context, this means defendant could be required to limit pollution say by 95 percent, paying damages for the harm caused by the remaining five percent pollution. In effect, the

injunction is structured so that it does not apply when the defendant's cost of compliance is disproportionately high, given the plaintiff's option to minimize damages by taking such simple precautions as the use of simple filters to control low-level pollution. That same strategy can work with patented technologies, so the ostensible hardship of patent injunctions can be effectively mitigated if the defendant is given time to make the corrections in the product, without being forced to recall infringing products that have been sold, or destroy infringing products made in good faith that have yet to be marketed.

This ability to mix and match remedies is a standard feature of equitable jurisdiction, yet it was wholly overlooked by Guido Calabresi and Douglas Melamed in their article, "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral."⁴⁰ Indeed their fundamental distinction between property and liability rules rests on the view that property rights allow the holder of an entitlement to keep it unless the outsider consents to meet his price. But the definition of a liability rule does not track common legal usage when it provides: "Whenever someone may destroy the initial entitlement if he is willing to pay an objectively determined value for it, an entitlement is protected by a liability rule."⁴¹ By this definition, liability is confined to cases that would be regarded as a deliberate harm—that is the force of the phrase "if he is willing to pay." This in contrast to common usage liability rules that attach as well to accidental destruction of property in part because it is simply too late to choose a form of property rights protection. Ordinary language thus distinguishes far more sharply between deliberate and accidental harms than does the Calabresi and Melamed model, and does so to good effect precisely to avoid the inefficient situation where one person is given a naked option to take the property of another for some unliquidated amount. The ability to integrate multiple remedies makes perfectly good sense if one thinks of a system of tort remedies as incremental, where the initial cut is the injunction against future deliberate harm, but it is modified incrementally as required.

There is, moreover, in this context no possibility of using the types of specific numbers that are key to contract remedies, but there nevertheless remains carryover between the tort and contract systems. Indeed the parallel is exceedingly close, because in both tort and contract the generalized duty of mitigation usually comes out second best in any situation where there is strict sequential performance, as took place in the construction contract cases noted above. The point is of great significance because it shows how it is possible to provide a unified framework that does not view any theory of incremental damages or efficient breach as the dominant model in civil litigation generally, or in litigation over patents in particular. It is important in all cases to tailor remedies to the particular problems faced. In general, it is wise to deal with consequential damages by limiting a plaintiff's recovery to induce plaintiff to take proper precautions post breach. But by the same token, it is unwise to limit those damages when the purpose of the legal system is to control the risk that a defendant will deliberately sidestep the voluntary contracting process on the one hand, or deliberately breach existing contracts, including existing licensing agreements on the other. It is for this reason that examination of contract and tort principles points to a system of strong injunctive relief to respond to the risk of misbehavior by parties who seek to use the property—whether real property, personal property, or standard essential patents created by others—whether real property, personal property, or standard essential patents—to their own advantage.

V. CONCLUSION

The most general conclusion that follows from this article is the fundamental unity between intellectual property and other forms of property. That unity manifests itself in dealing with contractual disputes between trading

parties on the one hand, and between strangers who could, but choose not to enter into voluntary arrangements on the other. In both settings, strategic behavior is a risk that must be countered. That risk takes the form of a hold-up between contracting parties and a hold-out with strangers. But in both settings the best legal rules recognize that the dangers of strategic behavior from deliberate breach rest with both parties, not just one. It is therefore unwise to tailor liability rules on the assumption that either a landowner or a patent owner is filled with guile, while an actual or potential trespasser or infringer acts with purely virtuous motivations.

The function of a sound remedial system is to counteract these tendencies. In those instances where consequential damages are at stake, low liquidated damages are most likely to induce optimal conduct on both sides. In this connection the failure of patented technologies give rise to the same problems as any other product failure. It makes no sense to say that the party who supplies a camera with a manufacturing defect

A STRONG PRESUMPTION IN FAVOR OF INJUNCTIVE RELIEF IS AN ESSENTIAL PART OF THAT SYSTEM, AS ARE DAMAGE RULES THAT MAKE IT COSTLY FOR PARTIES TO DEVIATE FROM VOLUNTARY NORMS OF COOPERATIVE BEHAVIOR.

must be held liable for the cost of reshooting the entire footage from a mountain climbing expedition. That same result holds when the failure results from a design defect in a patented product. In both cases, the looming risk is best solved by downstream adjustment whereby the product user takes along two or more cameras, often of different manufacture and design, to guard against that risk.

In contrast, in those contexts where the risk of deliberate breach by actual or potential licensees is significant, the situation changes. A strong presumption in favor of injunctive relief is an essential part of that system, as are damage rules that make it costly for parties to deviate from voluntary norms of cooperative behavior. In practice, the consequences of holdout or holdup are far smaller than the consequences of total disintegration of cooperative behavior, which flows uniquely from hold-out. The FTC makes the fundamental error of choosing the riskier path in its report. We are all well advised to resist its recommendations. ▲

1. Laurence A. Tisch Professor of Law, New York University School of Law, the Peter and Kirsten Bedford Senior Fellow, The Hoover Institution, and the James Parker Hall Distinguished Service Professor of Law Emeritus and Senior Lecturer, the University of Chicago. The research for this paper has been support by Qualcomm Inc.

2. Partner Cravath, Swaine & Moore L.L.P., former Under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office, 2009-2013.

3. For extensive commentary of the problem see Symposium, Property: A Bundle of Rights? 8 (3) Econ Journal Watch September 2011, available at <http://econjwatch.org/issues/volume-8-issue-3-september-2011>

4. Ronald H. Coase, The Problem of Social Cost, 3 J. L. Econ. 1 (1960). For subsequent work on the relationship of strong property rights, see Rafael La Porta, Andrei Shleifer, Florencio Lopez-Silanes & Robert Vishny, Law and Finance, 106 J. Pol. Econ. 1113 (1998)

5. For the history of these rules, see *Allnut v. Inglis*, 104 Eng. Rep. 206 (K.B. 1810), which worked its way into the American law of public utility regulation in *Munn v. Illinois*, 94 U.S. 113 (1876), from which develops the modern rules on rate regulation for public utilities in *Smyth v. Ames*, 169 U.S. 466 (1898) and *Federal Power Commission v. Hope Natural Gas*, 320 U.S. 591 (1944). For the doctrinal evolution, see Richard A. Epstein, *Principles for a Free Society: Reconciling Individual Liberty with the Common Good*, 282-287, 298-318 (1998).

6. On the importance of these issues, see Lon L. Fuller, *Consideration and Form* 41 COLUM. L. REV. 799, 800-801 (1941).
7. Lon L. Fuller, & William Perdue, The Reliance Interest in Contracts Damages, 46 Yale Law Journal 52, 373 (1936), available at, <http://www.cisg.law.pace.edu/cisg/biblio/fuller.html#vii> ; for a thorough modern critique see, Richard Craswell, *Against Fuller and Perdue*, 67 U. CHI. L. REV 99 (2000); Richard A. Epstein, *Beyond Foreseeability: Consequential Damages in the Law of Contract*, 18 J. LEGAL STUD. 105 (1989)

8. The Aristotelian influence is evident in this passage from Fuller and Perdue:

It is obvious that the three “interests” [expectation, reliance and restitution] we have distinguished do not present equal claims to judicial intervention. It may be assumed that ordinary standards of justice would regard the need for judicial intervention as decreasing in the order in which we have listed the three interests. The “restitution interest,” involving a combination of unjust impoverishment with unjust gain, presents the strongest case for relief. If, following Aristotle, we regard the purpose of justice as the maintenance of an equilibrium of goods among members of society, the restitution interest presents twice as strong a claim to judicial intervention as the reliance interest, since if A not only causes B to lose one unit but appropriates that unit to himself, the resulting discrepancy between A and B is not one unit but two.[6]

On the other hand, the promisee who has actually relied on the promise, even though he may not thereby have enriched the promisor, certainly presents a more pressing case for relief than the promisee who merely demands satisfaction for his disappointment in not getting what was promised him. In passing from compensation for change of position to compensation for loss of expectancy we pass, to use Aristotle’s terms again, from the realm of corrective justice to that of distributive justice.

9. Fuller & Perdue, 46 Yale L.J. at 55.
10. 547 U.S. 388 (2006).
11. This most useful formulation comes from Guido Calabresi, *The Costs of Accidents* (1970).
12. 286 N.W. 235 (Minn. 1939).
13. See, e.g., *Landis v. William Fannin Builders, Inc.* 951 N.E.2d 1078 (Ohio App 2011)(dispute over exterior of a new home).
14. U.C.C. § 2-712 (3) provides: ‘(3) Failure of the buyer to effect cover within this Section does not bar him from any other remedy.’ But that provision is immediately qualified in comment 3, which provides:
3. Subsection (3) expresses the policy that cover is not a mandatory remedy for the buyer. The buyer is always free to choose between cover and damages for non-delivery under the next section. However, this subsection must be read in conjunction with the section which limits the recovery of consequential damages to such as could not have been obviated by cover.
15. See, e.g., Charles J. Goetz & Robert E. Scott, *Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach*, 77 *Columbia Law Review* 554 (1977). For an early critique, see Daniel Friedmann, ‘The Efficient Breach Fallacy’ 18 *The Journal of Legal Studies* 1. (1989); Richard Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. Cal. L. Rev. 629 (1988).
16. Steven Shavell, *Damage Measures for Breach of Contract*, 11 *Bell J. Econ.* 466 (1980).
17. For a use of the single-owner framework, and the limitations thereon, see Richard A. Epstein, *Holdouts, Externalities and The Single Owner: Another Tribute to Ronald Coase*, 36 *J. LAW & ECON.* 553 (1993).
18. Oliver Wendell Holmes, Jr., *THE COMMON LAW* (Boston: Little, Brown and Co., 1881) 301. In fairness to Holmes, he also notes the role for equitable relief, which he considers to be only an “exceptional one,” *Id.* at 300-301, even though in many contexts injunctions and specific performance are often issued as a matter of course, i.e. as the presumptive remedy subject to exceptions, for example, as with laches.
19. See Uniform Commercial Code, § 2-304, following rules that date back to Roman times.
20. See U.C.C. § 2-204(3): (3) Even though one or more terms are left open a contract for sale does not fail for indefiniteness if the parties have intended to make a contract and there is a reasonably certain basis for giving an appropriate remedy.

21. See, Restatement Second of Torts § 357; Uniform Commercial Code, § 7-718.
22. See, e.g., Robert E. Scott & George Triantis, Embedded Options and the Case Against Compensation in Contract Law, 104 Colum. L. Rev. 1428 (2004).
23. Richard A. Epstein, F. Scott Kieff, and Daniel Spulber, The FTC, IP and SSOS: Government Hold-Up Replacement Private Coordination, 8(1) J. Competition L. & Econ. 1 (2012), <http://jcle.oxfordjournals.org/content/8/1/1.full?keytype=ref&ijkey=1tuKfhz2vhZ1CdU> . [hereafter EKS]
24. See, e.g., Alaska Packers' Ass'n v. Domenico 117 F. 99 (9th Cir. 1902).
25. EKS at 21.
26. No. C10-1823JLR (W.D. Wash. 4/25/2013)
27. 318 F. Supp. 1116 (S.D. N.Y.)
28. Minco, Inc. v. Combustion Eng'g, Inc. 95 F.3d 1109, 1119-20 (Fed. Cir. 1996)
29. 11-cv-09308 UPDATE
30. Innovatio at 8
31. For perceptive discussion, see Fennell v. Southern Maryland Hospital Center, Inc., 580 A. 2d 206, 209 (Md. 1990) See David Kaye, The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation, 1982 Am. B. Found. Res. J. 487
32. Steven Shavell, Economic Analysis of Accident Law 117 (1987)
33. Holderman slip at 19 (relying on Testimony of David Teece, and citing Certain Elec. Devices, Including Wireless commc'n Devices, Portable Music and Data Processing Devices, and Tablet Computers, Inv. No. 337-TA-794, at 63 (July 5, 2013) (Final).)
34. See infra at at v.
35. See infra at
36. Innovatio at 23., relying on Laserdynamics, Inc. v. Quanta Computer, Inc., 694 F.3d 51 (Fed. Cir. 2012)
37. Mark A. Lemley & Carl Shapiro, A Simple Approach to Setting Reasonable Royalties for Standard-Essential patents, SSRN http://papers.ssrn.com/so13/papers.cfm?abstract_id=2243026
38. EKS, at n. 24 . ALLIANCE FOR TELECOMM. INDUS. SOLUTIONS, COMMENTS ON P11-1204 at 1 (June 14,2011) ["ATIS Comments"] ("ATIS has not experienced the hold up problem"); INT'L COMM. FOR INFO. TECH. STANDARDS, COMMENTS ON P11-1204 at 1 (June 20, 2011)("The current officers and staff have not been notified of any active patent 'hold-up' problems with regards to INCITS standards."); TELECOMMS. INDUS. ASSOC., COMMENTS ON P11-1204 at 4 (June 14, 2011) ("TIA has never received any complaints regarding such 'patent hold-up' and does not agree that 'patent holdup' is plaguing the information and telecommunications technology (ICT) standard development processes.").
39. A. Mitchell Polinsky & Steven Shavell, The Uneasy Case for Product Liability, 123 Harv. L. Rev. 1437 (2010). The article notes the relevant dominance of market forces and direct regulation as a source of product safety. It does not, however, examine the evolution of product liability doctrine since the adoption of the Restatement, with its notable contraction of defenses based on plaintiff's conduct.
40. For application of this principle to both torts and takings law, see Richard A. Epstein, *The Takings Clause and Partial Interests in Land: On Sharp Boundaries and Continuous Distributions*, 78 BROOK. L. REV. 789 (2013); Richard A. Epstein, The Irrelevance of the Hand Formula: How Institutional Arrangements Structure Tort Liability" in Jef De Mot (ed.), *Liber amicorum Boudewijn Bouckaert*, Brugge, die Keure 65 (2012).
41. 85 Harv. L. Rev. 1089 (1972).
42. Id at 1092

Licensing Of Standard Essential Patents: Antitrust Intervention Is Not Big Enough A Fix

BY ELIANA GARCÉS TOLÓN¹

The recent antitrust interventions against patent holders issuing injunctions to assert standard essential patents have caused a stir and a debate on the role of antitrust enforcement in licensing negotiations. This piece argues that the way antitrust intervention is being framed allows regulators to restrict the behavior of the patent holder staying away from the issue of FRAND determination. Yet, uncertainty about the meaning of reasonableness and about what is and is not allowed under FRAND commitments lies behind most litigation activity surrounding SEPs. Without more clarity on what can or cannot be accepted under FRAND it is unlikely that substantial progress will be achieved in reducing costly litigation. Antitrust enforcement is ill placed to do the job. In the face of uncertainty about Courts' ability to develop a consistent line across the

THE WILLINGNESS OF ANTITRUST
AUTHORITIES TO ENGAGE IN
THE PREVENTION OF WHAT WAS
PERCEIVED AS A POSSIBLE HOLD-
UP BY OWNERS OF STANDARD
ESSENTIAL PATENTS PROVOKED
NEW INTEREST IN WHAT SHOULD
BE PERMISSIBLE IN THE ASSERTION
OF RIGHTS LINKED TO SEPS.... YET
THE SCOPE OF INTERVENTION
HAS BEEN VERY LIMITED AND
HAS FAILED TO ADDRESS THE
UNDERLYING SOURCE OF INEFFICIENT
LITIGATION: THE DISPUTE ABOUT
THE MEANING OF FAIR, REASONABLE
AND NON-DISCRIMINATORY
(FRAND) COMMITMENTS AND
THE SCOPE OF RIGHTS AND
OBLIGATIONS CONTRACTED WHEN
CONTRIBUTING TO A STANDARD.

globe, standard setting organizations have a role to play in shedding some light on fair licensing of SEPs.

The launch of investigations last year by antitrust authorities in the US and Europe relating to the recourse to injunction relief in the context of standard essential patents (SEPs) has generated a renewed interest in the role of antitrust intervention in such patent disputes. In particular, the adversarial relations between licensors and licensees of SEPs have generated questions on the actual implications of FRAND commitments encumbering these patents.²

Three cases were opened in 2012 by the European Commission. One related to a complaint by Apple that alleged an injunction sought by Samsung asserting standard essential patents amounted to anticompetitive behavior. The other two consisted of similar allegations against Google/Motorola by both Apple and Microsoft. Meanwhile, on the other side of the Atlantic, a settlement was reached between the Federal Trade Commission and Google after Google's acquisition of Motorola whereby Google committed to restrict its recourse to injunctions for infringements of its SEPs.

The willingness of antitrust authorities to engage in the prevention of what was perceived as a possible hold-up by owners of standard essential patents provoked new interest in what should be permissible in the assertion of rights linked to SEPs. As a consequence, discussions on these issues started and are still continuing in two major international standard-setting organizations (SSOs), ETSI and ITU. The divergence of interests among

members of these organizations and the diverse interpretations given to FRAND related rights and obligations have so far prevented an agreement on how the SSOs policies could be made more precise and more conducive to an efficient standard setting process.

The request for intervention by antitrust regulators was a development in the mounting litigation battles concerning the licensing of patents between major ITC players. Patent disputes, including disputes surrounding SEPs, have traditionally been brought to courts under patent law. But some companies brought up the argument that standard essential patents were not just being asserted in order to obtain remuneration, but were being abused for the purpose of gaining illegitimate advantage in the market. Courts were not efficiently resolving these issues because they were mostly concerned with establishing patent infringement without taking into account the damage that the litigation process itself, often unmeritorious, was causing to competition. In addition, some in Europe considered certain procedures, such as the one established by the Orange Book case in Germany, imposed excessive obligations on potential licensees.³ Under Orange Book, an implementor faced with an injunction can defend itself against the patent holder but must irrevocably commit to take a license on FRAND terms and deposit a reasonable remuneration in escrow. This precludes any defense based on the merits of the patent assertion.

It was on the basis of such arguments that antitrust enforcers felt legitimized to intervene. Yet the scope of intervention has been very limited and has failed to address the underlying source of inefficient litigation: the dispute about the meaning of fair, reasonable and non-discriminatory (FRAND) commitments and the scope of rights and obligations contracted when contributing to a standard.

I. THE BASIS FOR ANTITRUST INVOLVEMENT IN SEP DISPUTES

Antitrust enforcement only rarely intervenes against the enforcement of intellectual property rights by patent owners, and almost never against refusal to license to direct competitors in the market where the patent is implemented. Refusals to license are deemed to be anticompetitive only if they relate to intellectual property rights that are objectively necessary to compete in a new downstream market that risks being eliminated without the license. In addition, the refusal to license must lead to consumer harm and not be objectively justified.⁴

In the case of SEPs, antitrust enforcers on both sides of the Atlantic felt they needed to intervene against the recourse to injunctions by firms seeking to assert their SEPs against implementers. Two facts about SEPs make them susceptible to antitrust intervention. First, the fact that SEP reading in a standard may become de facto essential for players wanting to enter the market where the functionality of the standard is used. This can create opportunities for hold up by the patent owner. Second, SEPs are more often than not encumbered by FRAND commitments designed to provide some guarantee of fair access to licensees. This is in fact a policy established by standard setting organizations aimed at mitigating the hold-up risk. We further analyze the basis for antitrust intervention and explore its limitations.

A. Hold-Up in SEP Disputes

Antitrust intervention in SEP disputes has been justified as a remedy against hold-up.⁵ In a hold-up, the owner of a standard essential patent attempts to use the market power obtained through the commitment to the technology by the licensee to extract more generous license terms than those that it could obtain if it were competing with

other technologies. Such hold-up is possible when future licensees make sunk investments that commit them to the technology before they actually negotiate the license fees. In these cases, it is possible for the licensor of an essential patent to extract some of the value of such sunk investment by raising the value of licensing terms.⁶ The licensee will be willing to pay some additional license fees rather than to incur the more expensive costs of

BUT IN THEIR RECENT INTERVENTIONS, EU ANTITRUST REGULATORS HAVE INVOKED A VIOLATION OF ANTITRUST RULES TO INVESTIGATE WHAT WERE CONSIDERED INSTANCES OF HOLD-UP BY SEP OWNERS. THE INTERVENTIONS HAVE CENTERED ON THE REQUESTS OF INJUNCTIONS BY SEP LICENSORS AGAINST IMPLEMENTERS.

switching technologies. In the context of standards, implementers often make investments for the adoption of the standard before they conclude negotiations for licensing terms. This makes them susceptible to hold-up by SEP owners.

Antitrust intervention has not yet sought to directly address cases in which SEP owners have attempted to extract illegitimate rates from patentees by way of hold-up. In 2008, the FTC condemned a company, Negotiated Data Solutions LLC, for attempting to

unilaterally raise its pre-agreed license fees to a number of clients after the clients had adopted and invested in the technology.⁷ But the case was brought under Section 5 of the FTC Act prohibiting unfair or deceptive practices affecting commerce, normally used under consumer protection. It was not argued as an antitrust violation under the Sherman Act, but rather the conduct was deemed “inherently coercive and oppressive” given the lock-in of the implementers.

But in their recent interventions, EU antitrust regulators have invoked a violation of antitrust rules to investigate what were considered instances of hold-up by SEP owners. The interventions have centered on the requests of injunctions by SEP licensors against implementers. While in the US the FTC complaint in the matter of Motorola Mobility was also based on a violation of Section 5 of the FTC Act, in the EU the cases relating to requests for injunctions against SEP implementers were treated as violation under Article 102 TFEU, condemning abuses of market power.

In order to construct a hold-up case as an Article 102 antitrust violation, one needs to argue that a request for high royalties to an implementer that has invested in the technology can constitute an abuse of market power. Classifying this as an abuse of market power establishes that the market power derived from the pre-commitment of the licensee to the standard is illegitimate and should not be exercised. One argument in support of this is that the additional rent that a patent owner can extract from implementers because these have already committed to the standard is not correlated to any additional technological value of the patent and is therefore not legitimate. This argument relies on the fact that the inclusion of the patent in a standard is not an innovation nor does it increase the actual social value of the technology.⁸ Therefore there is no justification for adjusting the value of the patent to account for the market power generated by the participation in the standard, and the SEP owner should be deterred from doing so. Under this argument a hold-up can be construed as the exercise of illegitimate market power and an antitrust violation.

B. Injunctions as a Per-se Hold-Up

The ability to request an injunction against a SEP licensee during licensing negotiations can provide with a powerful mean to extract onerous royalty terms. The cost to the implementer of having a product withdrawn

from the market can be enormous, and the risk of such an event happening will increase substantially the price that a licensee is willing to pay. At the same time, because the SEP holder has normally committed to license its SEP on FRAND terms and, in consequence, a refusal to license is not an option, the recourse to injunctive relief can then only be interpreted as a strategy to increase the value of the licensing terms. Because the pressure on the licensee brought by the possibility of an injunction is likely to result in higher value licensing terms, a request for an injunction-asserting SEP after the standard is adopted can be considered a way of achieving a hold-up. Such recourses to injunctions are therefore interpreted as evidence that illegitimate market power is being exercised.

Interestingly, it is the negotiating behavior and not the actual negotiation outcome that points to the possibility of hold-up and anticompetitive conduct. Under this reasoning there is no need to define the boundaries of FRAND licenses because the harm is assumed from the process of negotiation. An injunction requested by the SEP owner is assumed *per se* to lead to illegitimate additional rent.

This may not technically represent a departure from an effects-based assessment in that the reasoning behind the finding of an abuse rests on a very likely negative impact on the licensing fee. But it obviates the need to actually assess the actual impact of the behavior. As a consequence, there is no need to evaluate whether the outcome of a negotiation under the threat of an injunctive relief falls within the boundaries of FRAND terms. In fact, that question does not seem, in a pure hold-up case, to be relevant.

II. CAN ANTITRUST HELP IN FRAND DETERMINATION?

In the opening of the investigation against Samsung, the European Commission stated that it would investigate whether Samsung's recourse to injunctive relief amounted to a breach of FRAND commitments. But the reality is that demonstrating such violation does not seem material to the case if the case is argued as a hold up case.

THE ABILITY TO REQUEST AN INJUNCTION AGAINST A SEP LICENSEE DURING LICENSING NEGOTIATIONS CAN PROVIDE WITH A POWERFUL MEAN TO EXTRACT ONEROUS ROYALTY TERMS. THE COST TO THE IMPLEMENTER OF HAVING A PRODUCT WITHDRAWN FROM THE MARKET CAN BE ENORMOUS, AND THE RISK OF SUCH AN EVENT HAPPENING WILL INCREASE SUBSTANTIALLY THE PRICE THAT A LICENSEE IS WILLING TO PAY.

A. *Hold-up Theory does not help with FRAND determination*

Antitrust intervention against injunctions to assert SEPs relies on the hold-up theory rather than on a violation of FRAND terms. None of the reasoning justifying antitrust intervention based on hold-up theory seems to rely on the existence of FRAND commitments, even though the existence of such commitments is sometimes referred to as further evidence of abuse. Once established that the extraction of market gains from participation in the standard is illegitimate and unwarranted because it does not reflect the value of the technology for the licensee, showing a violation of FRAND terms is superfluous. It is at best an exercise of tautology where the violation of FRAND and the definition of the infringement are done concomitantly.

By definition, FRAND terms will be violated if there is an attempt to extract illegitimate rent from the

licensee, and in this sense antitrust intervention informs on the definition of FRAND by making explicit the kind of rent it should not include and the kind of behavior that would be suspicious of extracting that rent opportunity. But so far the investigations opened have stayed away from defining the value or terms of the license that is in fact legitimate to extract. Many will welcome this as prudent behavior and argue that it is not for antitrust regulators to define fair, reasonable and non-discriminatory licensing terms.⁹

NONE OF THE REASONING JUSTIFYING ANTITRUST INTERVENTION BASED ON HOLD-UP THEORY SEEMS TO RELY ON THE EXISTENCE OF FRAND COMMITMENTS, EVEN THOUGH THE EXISTENCE OF SUCH COMMITMENTS IS SOMETIMES REFERRED TO AS FURTHER EVIDENCE OF ABUSE.

Antitrust intervention in Europe seems so far to have been quite careful not to be seen as determining the scope of FRAND commitments or finding a violation of FRAND terms. Although it has contributed to the determination of FRAND commitments by arguing that the recourse to injunction is inadmissible in some cases, it has only done so indirectly by restricting the patent owner's behavior during the negotiating process. Beyond this behavioral restriction, there have

been no limitations by antitrust regulators so far of what FRAND terms can or cannot include.

One can only relate the lack of enthusiasm by antitrust authorities to get involved in the discussions surrounding the meaning of FRAND to the general avoidance in past years of cases involving arguments of excessive pricing and exploitative abuses. This aversion to getting involved in anything that could be associated with price determination is in fact the main reason why antitrust intervention is not well equipped to resolve the litigious dynamics that have developed around SEP licensing.

B. FRAND determination requires a theory of fair value

The general reticence so far to determine more concretely what is and what is not FRAND is not justified by the lack of analytical tools. A theoretical body of literature has addressed the question on how to value the technology absent the market power of the standard. The most pervasive methodology proposes a valuation based on the ex-ante value of the technology, that is, before the technology is included in the standard.¹⁰ Ways to calculate such valuation include estimating the value of the differential impact of a technology compared to the next best alternative. Such methodologies, if implemented, could help determine the “excess price” that is gained by hold-up. This ex-ante methodology has been endorsed by the European Commission as one acceptable way to approach FRAND.¹¹ But no methodology has yet been proposed in an antitrust investigation, and antitrust regulators have stopped short of defining any range for FRAND terms.

One could raise the valid argument that the current licensing terms agreed between licensing parties involve a complex equilibria of reciprocal commitments and that this does not make FRAND terms susceptible to optimal ex-ante determination. Licensing contracts for SEPs include long lists of specific rights and obligations by the signing parties and such agreements are best left to be negotiated without any outside constraints. How far can one go with determining FRAND by behavior? If such is the case, the approach adopted by antitrust authorities in Europe to only sanction behavior that is susceptible to distort negotiations is correct. The question is then whether there are other means than injunctions to extract illegitimate rent. So far, no other theory of harm has been taken up by regulators in the EU or the US, although some players in the industry have raised concerns about behavior such as the bundling of patents or the cross-licensing demands of certain SEP hold-

ers. Assessing whether such demands are consistent with FRAND or not is in fact very difficult to do without addressing upfront the question of what constitute reasonable and fair licensing terms in the case of SEPs. Antitrust enforcers have so far shown no inclination to do this.

III. WHERE TO LOOK FOR PRINCIPLES OF FRAND DETERMINATION?

And yet the lack of clarity of the meaning of FRAND continues to generate costly and possibly inefficient litigation. There is no definition of FRAND in any of the standard setting organizations that require it as a commitment for licensing. Current attempts by the industry to find common definitions, and even common principles, have brought to light the extent of disagreement among market participants. There is not one single aspect surrounding the definition of FRAND that today generates unanimity. Firms have been turning to courts to resolve licensing disputes, and the latest attempt to turn to antitrust law will shed little more light on the issue. In this context, one might want to turn to the standard setting organizations for a more efficient way to provide clarity on the boundaries of FRAND.

A. *The current disputes in FRAND*

Disputes surrounding FRAND touch upon all possible aspects of the concept. At present there is no agreement among industry participants of whether FRAND terms should be tilted towards guaranteeing incentives to innovate by patent holders, or whether they should aim at a socially optimal dissemination of the technology. Beyond such general principle, the more detailed application of FRAND principles is even more divisive.

Disputes and divergence of interpretation also concern the determination of the value of a technology including for example the product for which the value of the patent must be calculated. Should the base be the value of the final product that has the standard embedded, or should the base be the smallest sellable component of the product containing the standard? Other disputes relate to whether FRAND commitment should apply patent-by-patent, or to the entire standard in which cases the stacking of license fees needs to be taken into account. Controversy goes further. Are licensees of SEPs entitled to patent-by-patent licensing, or are FRAND licensing commitments compatible with imposing portfolio licensing? Are the FRAND commitments transferred with the SEPs when they are sold, or do these commitments only bind the original owner who made the patent declaration? Are cross-licensing demands compatible with FRAND, or should there always be a monetary option for the patentee? Are cross licensing of standard essential patents reading to the same standard legitimate in FRAND? What about cross licensing of all SEP or cross licensing of non-SEPs?

...THE LACK OF CLARITY OF THE MEANING OF FRAND CONTINUES TO GENERATE COSTLY AND POSSIBLY INEFFICIENT LITIGATION.... THERE IS NOT ONE SINGLE ASPECT SURROUNDING THE DEFINITION OF FRAND THAT TODAY GENERATES UNANIMITY. FIRMS HAVE BEEN TURNING TO COURTS TO RESOLVE LICENSING DISPUTES, AND THE LATEST ATTEMPT TO TURN TO ANTITRUST LAW WILL SHED LITTLE MORE LIGHT ON THE ISSUE.

These questions are not merely rhetorical but the subject of expensive litigation activity. According to a study commissioned by the European Commission, SEPs are five times more likely to be litigated than non-SEPs of

similar characteristics, and the propensity for litigation has increased steadily in the last 20 years.¹² Yet there are no answers to be found in antitrust enforcement, nor in the policies or declaration forms of any existing standard setting organization. But clarifying some FRAND rules could decrease the amount of costly litigation and produce a more efficient system of standard formation and standard licensing by providing legal certainty.

As it was argued before, it seems unlikely that antitrust cases will venture into this territory in the near future, but it might give enough theoretical ammunition to support the idea that rules that limit the possibility or the attractiveness of hold-up should be favored. Companies are currently turning to courts to resolve disputes and, at least in the US, they seem to be rising to the task.

B. *The Role Of Courts*

Most licensing disputes settle without any need for a definition of FRAND terms. In those cases that are litigated, courts have mostly established infringement and declared damage claims without making any attempts to define the terms of FRAND licenses. The ruling in *Microsoft v. Motorola* in 2013 marked a departure and an important precedent. In that case the US District Court judge set out to determine what he considered should be the fee of the FRAND license in that case.¹³ Judge James L. Robart established a methodology that embraced the principle of linking FRAND to the ex-ante value of the technology and based its determination on the license fee that the patent owner could have obtained from a negotiation absent the effect of the standard. In essence, Judge Robart endorsed the principle promoted by antitrust regulators that the market power derived from the participation in the standard is illegitimate. As a practical solution, Judge Robart used the price of a similar technology in a patent pool as a benchmark for undistorted negotiations. He took into account the relative importance of that particular patent to the standard, as well as the importance of the patent to the final product of the licensor.

Thus, in its findings, the Court seemed to go further than just establishing the principles supported by antitrust regulators, and provided much more clarity on the meaning and value of FRAND terms. First, Judge Robart established that the objective of FRAND terms is not only to induce participation in a standard but also to facilitate the dissemination of the technology. Second, the fact finding explicitly said that royalty stacking is an issue to be taken into account in the FRAND determination of a single patent.

In September 2013, another US judge followed a similar reasoning in the lawsuit initiated by patent holder Innovatio IP Ventures. In this case, Judge James F. Holderman took the smallest sellable unit in which the patent and the standard were used as the royalty base for the FRAND calculation.¹⁴

In both cases, the determined FRAND licensing fee was significantly lower than the one sought by the patent holder.

So far, it seems that the reasoning used by courts in the United States for the determination of FRAND is consistent with the “hold-up” theory supported by antitrust regulators. But judges in the US seem ready to go further than antitrust theory ventures to go by establishing such concepts as the purpose of FRAND terms, the elements to be taken into account and even the proper base for the value of the patent.

On both sides of the Atlantic, antitrust regulators are proposing remedies to the issue of injunctions requests

to assert SEPs that involve FRAND determination by a Court or an arbitration body. This means that courts or arbitrators will possibly play an increasingly important role in determining the definition and possible range of FRAND terms. Courts in the US have so far taken the lead in providing some clarity on the meaning of FRAND terms, but it remains to be seen whether, and to what extent, EU courts will follow and whether consistency will be achieved.

C. The Role Of Standard Setting Organizations

The standard setting organizations have so far mostly stayed away from the issue of the determination of FRAND. But they could possibly play a bigger role in promoting efficient rules and improving the environment for standard setting. It is hard to argue at this point that they take no part in improving the licensing environment of standard essential patents.

Because SSOs are governed by their members, it is unlikely that SSOs will provide a precise definition of what FRAND terms are. Negotiated licensing arrangements can be very complex, and SSO members are unlikely to want to excessively restrict the flexibility of such contracts. Yet SSOs could take some useful steps that would provide legal clarity and reduce litigation. First, they could take measures that render hold-up more difficult. SSOs could include a general statement supporting the principle that the exercise of market power derived from participation in the standard is illegitimate even though the practicality of such statement is unclear. More actionable measures would be to explicitly link the FRAND commitment to the patent and not the owner so that the commitment survives transfer of ownership. Promoting ex-ante declaration of maximum licensing terms is another example of a way to decrease the risk of hold-up. Such ex-ante declaration has already been promoted by SSOs such as IEEE and VITA.¹⁵

In addition to imposing some restrictions in licensing behavior, SSOs would greatly improve the patenting environment in standards if they put some effort into improving the patent declaration procedures, notably through a better assessment of validity and essentiality of the patents declared. Although most patent litigation in the context of SEPs relates to a dispute in licensing terms, the pervasive over-declaration of patents in standards has incentivized players to challenge the actual validity of the patents subject to negotiations disputes. Measures to increase the quality of the patent declarations would reduce incentives for litigation by increasing the likelihood of patent validity and essentiality.

The provision of a system of arbitration would also limit the strategic use of litigation and accelerate the resolution of disputes. Such arbitration rules would have to be designed to be attractive to all members.

It is far from easy to obtain consensus for such solutions within SSOs given the divergent interest of the members. Yet many regulators have supported a greater involvement of SSOs in the solution to the current high

SO FAR, IT SEEMS THAT THE REASONING USED BY COURTS IN THE UNITED STATES FOR THE DETERMINATION OF FRAND IS CONSISTENT WITH THE “HOLD-UP” THEORY SUPPORTED BY ANTITRUST REGULATORS. BUT JUDGES IN THE US SEEM READY TO GO FURTHER THAN ANTITRUST THEORY VENTURES TO GO BY ESTABLISHING SUCH CONCEPTS AS THE PURPOSE OF FRAND TERMS, THE ELEMENTS TO BE TAKEN INTO ACCOUNT AND EVEN THE PROPER BASE FOR THE VALUE OF THE PATENT.

litigation environment.¹⁶ Any reform will have to ensure all relevant industry participants stay on board. The European Commission is currently actively engaging with SSOs in Europe in order to explore the possibility for progress. It is at this point too soon to predict what will be achieved in the current context.

IV. CONCLUSION

SO FAR, IT SEEMS THAT THE REASONING USED BY COURTS IN THE UNITED STATES FOR THE DETERMINATION OF FRAND IS CONSISTENT WITH THE “HOLD-UP” THEORY SUPPORTED BY ANTITRUST REGULATORS. BUT JUDGES IN THE US SEEM READY TO GO FURTHER THAN ANTITRUST THEORY VENTURES TO GO BY ESTABLISHING SUCH CONCEPTS AS THE PURPOSE OF FRAND TERMS, THE ELEMENTS TO BE TAKEN INTO ACCOUNT AND EVEN THE PROPER BASE FOR THE VALUE OF THE PATENT.

European regulators remain committed to European-wide standards and consider standardization processes as part of its internal market and industrial strategy. For this they have reacted with some concern for the increase of litigation and disputes in the context of IPR encumbered standards. The recourse to antitrust intervention is likely to result in some boundaries on the behavior of SEP licensors and, more precisely, in a restriction of their ability to request injunctions. But this is not enough. More clarity is needed about the actual boundaries of FRAND terms and antitrust policy is, for the moment, unlikely to intervene further. Courts in the US have started to provide some clarity on the meaning and calculation of FRAND terms, but in Europe courts are still hesitant on the issue. SSOs are well placed to improve the predictability of patent

licenses by improving the quality of declaration and by adopting some measures aimed at reducing the opportunity for hold-up. It will be a challenge to bring the industry together on this project, but it is a necessary effort. ▲

1. Eliana Garcés Tolon is a Senior Economist and Deputy Head of the Unit for Industrial Competitiveness Policy at the European Commission. She is a former member of cabinet of Vice President Almunia, European Commissioner for Competition Policy. The comments presented in this article reflect her personal opinion and by no means represent an official position of the European Commission.

2. FRAND stands for Fair, Reasonable and Non Discriminatory. These licensing commitments are commonly called RAND in the US.

3. German Federal Supreme Court, KZR 39/06, «Orange Book», May 2009

4. Commission Decision 89/205/EEC of 21/12/1988 *Magill TV Guide/ITP*, BBC, RTE OJ L 78

5. See FTC complaint in the matter of Motorola Mobility LLC and Google Inc., File No. 121-0120

6. A theoretical explanation of hold up-in SEPs can be found in Farrell, J., Hayes, J., Shapiro, C., & Sullivan, T. (2007). Standard setting, patents, and hold-up. *Antitrust Law Journal*, 74(3), 603-670.

7. *N-Data*, 2008 WL 258308,

8. Farrell, J., Hayes, J., Shapiro, C., & Sullivan, T. (2007). Standard setting, patents, and hold-up. *Antitrust Law Journal*, 74(3), 603-670.

9. See Mariniello, M. (2012) 'Standard-setting abuse: the case for antitrust control' Bruegel Policy Brief - February 2012

10. cf. 6; Swanson, D. G., & Baumol, W. J. (2005). Reasonable and nondiscriminatory (RAND) royalties, standards selection, and control of market power. *Antitrust Law Journal*, 73(1), 1-58.

11. Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-

operation agreements - Official Journal C11, 14.1.2011

12. 'Patents in Standards: a modern framework for IPR based standardisation' ECSIP consortium for the European Commission (forthcoming)
13. Microsoft v. Motorola, 2013 WL 2111217 (W.D. Wash., Apr. 25, 2013)
14. Innovatio IP Ventures Patent Litigation, MDL 2303 (N.D. Ill., Sept. 27, 2013)
15. Patent Challenges for Standard-Setting in the Global Economy, Keith Maskus and Stephen Merrill Eds. National Research Council. 2013.
16. Kuhn, K.U., Scott Morton, F. and Shelanski, H. (2013) 'Standard Setting Organisations can help solve the Standard Essential Patent licensing problem'. CPI Antitrust Chronicle, March 2013; Hesse, R (2012) 'Six Small Proposals for SSOs before Lunch' Remarks as Prepared for the ITU-T Patent Roundtable. U.S. Department of Justice.

Justifying Antitrust Intervention in ICT Sector Patent Disputes: How to Address the Hold-Up Problem

BY KAI-UWE KÜHN¹

The economic justification for any regulatory intervention in patent litigation, especially those for standard essential patents, comes from the view that hold-up of users of patents is endemic to some industries, especially ICT. The paper reviews these reasons why hold-up is more likely in ICT industries and discusses the type of evidence that is available. It discusses the strengths and limitations of competition policy instruments and notes that addressing the issue is far more appropriate under an abuse of dominance standard that allows for exploitative abuses to be addressed. It is finally explained that nevertheless the use of such an instrument has severe limitations and explores the question of how incentives of standard setting organizations can be improved to make commitments to FRAND licensing more meaningful.

I. INTRODUCTION

With escalating patent disputes across the world and threatened and achieved injunctions in some jurisdictions (as well as the International Trade Commission in the US), there is a sense that the patent system is in crisis.² This is in particular the case in the ICT industries in which patent litigation has escalated over the past decade and the large multiplicity of patents related to any device appears to lead to a significant likelihood that any dispute concerning a small feature of the product can lead to temporary exclusion of socially valuable products

WITH ESCALATING PATENT DISPUTES ACROSS THE WORLD AND THREATENED AND ACHIEVED INJUNCTIONS IN SOME JURISDICTIONS (AS WELL AS THE INTERNATIONAL TRADE COMMISSION IN THE US), THERE IS A SENSE THAT THE PATENT SYSTEM IS IN CRISIS.² THIS IS IN PARTICULAR THE CASE IN THE ICT INDUSTRIES

from the market. In addition, it is claimed that such escalating litigation and the associated liability risks are bound to chill innovation—at least in fast moving markets. In fact, there is even a strong view on part of some economists that a complete abolishment of patents would be better for innovation and efficiency in the economy.³ They point to the consistent evidence that has emerged in recent years that competition is a strong positive driver of innovation and claim that first mover advantages are sufficient to give innovation incentives in fast moving markets.

In this paper, I argue the case for policy intervention in the patent disputes in the ICT sector from a more traditional view that takes the proposition that patents are a valuable incentive device for inducing innovation as given. This means that some distortion to optimal ex-post production is seen as a necessary second best device for giving incentives to innovation. However, such an incentive system has to strive to minimize the distortion that is created in consumption.

I will discuss in this paper that the most serious distortion that can arise is what economists call the hold-up problem. I explain the economics of the hold-up problem and why it is particularly severe in the ICT industries. These reasons include the fact that patent boundaries are almost impossible to determine ex-ante—even with a

good patent examining system.⁴ This means that patent claims are necessarily quite uncertain, distinguishing intellectual property rights from other property rights. Furthermore, the large multiplicity of potential patents for any product offered in the ICT industries augment the uncertainty about infringement and validity for anyone considering investments into developing a new product. This combination of features also makes it impossible to write complete contracts about licensing conditions for existing patents before investing into the development of a new product

At the same time the costs of designing around a feature that has been incorporated into a product can be very high—especially when the product is forced off the market during the litigation period through preliminary injunctions.⁵ Together with the inability of the developer of a new product to write *ex-ante* contracts for all relevant patents, the fact that workarounds are costly, makes hold-up of new product developers by owners of previously existing patents endemic to the patent system.

I explain why the issue of hold-up is particularly clear in the case of standard essential patents (SEPs) and why the problem of hold-up will lead to more severe distortions when industry participants are very asymmetric in their intellectual property portfolios. Such asymmetries are particularly pertinent when evaluating the role of non-practicing entities.

Before evaluating the policy instruments that may be available to alleviate the hold-up problem, I will address the issue whether empirically hold-up actually exists. It has recently been claimed that there is no evidence that there is a hold-up problem and therefore intervention is not justified. I argue that this claim arises from a misunderstanding of the nature of economic evidence and is fundamentally mistaken in its approach to evaluating counterfactuals. I will show that there is a solid body of theory and indirect evidence supporting the underlying assumptions of the theory (including indirect evidence on the importance of hold-up) that is as strong a positive economic case as one can make for intervention if one has a realistic view of the evidence that can be available. Secondly, there is little evidence that proposed interventions would have any negative effect on the incentives to innovate in the ICT industries. In the context of imperfect evidence, the weighing of likely positive effects of intervention and likely harm does indeed lead to a strong justification for policy intervention.

A more difficult question is what the right policy tools are for intervention aimed at correcting distortions caused by hold-up. I discuss these issues first in the context of SEPs. I note that in this context it appears at first surprising that the hold-up issues cannot be resolved by contracts. However, empirical evidence over a significant period of time shows that there seem to be almost insurmountable barriers to achieving such contracting solutions in the context of standard setting organizations without governmental intervention. I discuss the mixed record of court systems in the US and Europe and the limitations to resolving hold-up issues through the political process.

On this background I discuss the scope for antitrust intervention. I argue that the abuse of dominance standard is a more natural fit for addressing the hold-up problem than the monopolization standard. In particular, actions that clearly lead to a worsening of the hold-up problem can be directly identified as an abuse of a dominant position, while demonstrating foreclosure is not only much harder but also does not fit very well with the fundamental underlying issue of hold-up. I explore the limits of antitrust enforcement against individual firms and the scope for antitrust rules that would impose constraints on standard setting organizations.

II. THE PERVASIVENESS OF HOLD-UP AND ITS IMPORTANCE IN ICT INDUSTRIES

A. *The Basic Hold-Up Problem*

In well-functioning markets the creator of a new product would incorporate features into the product taking fully account of the license fees he has to pay for such incorporation. There would be competition between different solutions and the creator of the product would simply choose a solution that gives the best trade-off between licensing costs and value added to the overall product. In such an ideal world contracts are complete. If patents were to convey a monopoly the inventor would at least bargain the payment down to a point in which the feature would just be included in the product. The owner of the product has an incentive not to set excessive royalties in order to create an incentive for innovations to be adopted that include his intellectual property.⁶

However, in many real life situations the designer of a product has to make investment decisions that lock him into a single solution before he can negotiate with the patent holder whose rights are affected. To start with the most extreme case, suppose for the moment that after the investment into product development has been made, there is no possibility at all to sell the product without infringing on a patent. If a license could not be contracted on ex-ante, the investment into creating the new product is then sunk and the producer of the new product will bargain over the whole value of the product with the patent holder. This can be a much higher value than just the value of the product net of investment costs, which would be the case with ex-ante negotiation. The result is that any new product that is reliant on a specific patent to be able to stay in the market will make less of a return on investment and the incentives to innovate on products that might infringe on existing intellectual property as an input will decline.

The above argument assumed that the firm with the new product cannot avoid the patent by redesigning the product. In the real world there are often such possibilities. However, hold-up can still be severe. The reason is that it will always take some time to redesign a product to avoid infringement. The scope for hold-up then depends on the costs that can be imposed on the firm that is trying to sell the new product during the period of time in which it can redesign the product. If redesign is costless there is no hold-up. However, there are few situations in which redesign are of low cost. There are always delays to implement a new design even when the features are already known. The losses in the market place induced by not being able to sell for even a short period of time can be large. Take as an example the litigation between NTP and RIM between 2000 and 2006. Under threat of an injunction, which potentially could have interrupted Blackberry service for a short period of time, RIM settled for over \$600 million. This sum most likely primarily reflected the massive long run expected loss of business arising from a spectacular shut down of service. Note that the market value of a patent in such circumstances does not reflect the intrinsic increase in value of the product that the patent generates but instead the value of the potential *costs* that can be induced through an injunction. The fact that market valuations seemed to value the patents even higher than the eventual settlement amount is therefore no evidence against hold-up. The fact that several of the relevant patents were later invalidated by the US patent office is rather evidence in the opposite direction. But by the time of patent invalidation, it was too late for RIM because the potential costs of delaying a settlement were just too high.

Similar ways of inducing large losses on a generation of devices by threatening and/or enforcing an injunction can be seen in the current patent wars. For example, Apple achieved an injunction in August 2011 (confirmed in

an appeals decision in September 2011) against Samsung's Galaxy 10.1 tablet in Germany. Although Samsung developed a workaround in only a month, it took until January 2012 to lift the sales ban. In this way Samsung missed a whole Christmas season for one generation of its devices despite being able to present a workaround in October. What this shows is that there are enough frictions in the system that can lead to very substantial losses even when workarounds are of low cost. Hence, hold-up potential is clearly high.

B. Why Hold-Up Problems Are More Severe in ICT Industries

Are all of the hold-up issues discussed above not simply the fault of RIM or Samsung? Shouldn't they have looked at the patents that they might have infringed and made a non-infringing product in the first place? Alternatively, they could have just licensed the patents they might infringe with the new product. If these strategies were available, there should be little sympathy for a firm that gambled by not acquiring a license. Why should they not pay heavily? They took the risk of entering the market without a license instead of negotiating for it beforehand.

What such an assessment claims is that there is never any hold-up problem. It proposes that any contracting issue over patent licenses can be resolved before an investment into a new product takes place. Such an approach to the analysis of the current problems in IP licensing in the ICT industries overlooks that there is no perfect market for intellectual property. Indeed, some of the imperfections in the intellectual property system are particularly severe in the ICT industries. First, the boundaries of intellectual property rights are particularly unclear in this area.⁷ This is not a matter just of inefficiently working patent offices. An early example in a high tech industry that illustrates this point is the dispute between Polaroid and Kodak after 1976. Kodak had extensively researched and assessed the Polaroid intellectual property rights on instant photography and concluded on the basis of expert reports that there was no infringement. And although the judge in the case conceded that Kodak had done everything to avoid infringement (and invent around Polaroid's intellectual property rights), it was found to infringe nevertheless.

The second issue is the proliferation of complementary patents (of uncertain validity and uncertain boundaries) in an industry that is particularly fast moving. The number of patents that could potentially be infringed by a new product is very large. A modern smart phone contains parts covered by thousands of patents—most of them uncertain in their scope and validity. Many potentially relevant patents will not be known to the designer of the

...WAYS OF INDUCING LARGE LOSSES ON A GENERATION OF DEVICES BY THREATENING AND/OR ENFORCING AN INJUNCTION CAN BE SEEN IN THE CURRENT PATENT WARS. FOR EXAMPLE, APPLE ACHIEVED AN INJUNCTION IN AUGUST 2011 (CONFIRMED IN AN APPEALS DECISION IN SEPTEMBER 2011) AGAINST SAMSUNG'S GALAXY 10.1 TABLET IN GERMANY. ALTHOUGH SAMSUNG DEVELOPED A WORKAROUND IN ONLY A MONTH, IT TOOK UNTIL JANUARY 2012 TO LIFT THE SALES BAN. IN THIS WAY SAMSUNG MISSED A WHOLE CHRISTMAS SEASON FOR ONE GENERATION OF ITS DEVICES DESPITE BEING ABLE TO PRESENT A WORKAROUND IN OCTOBER. WHAT THIS SHOWS IS THAT THERE ARE ENOUGH FRICTIONS IN THE SYSTEM THAT CAN LEAD TO VERY SUBSTANTIAL LOSSES EVEN WHEN WORKAROUNDS ARE OF LOW COST. HENCE, HOLD-UP POTENTIAL IS CLEARLY HIGH.

new product. Indeed, by coming up with the new product, the designer would have independently discovered the invention, if a patent is actually infringed. Kodak, for example, spent 10 years from the development of its technology to market entry trying to avoid patent infringement and failed. With the current innovation cycle in the ICT industry such careful and time consuming patent search and assessment would not be possible if one would want to compete with any product in the market. As a result, firms will have to invest in new products and develop them, knowing that there will likely be some infringement but that they cannot tell beforehand which part of the product is likely to infringe a patent and who holds the relevant intellectual property. This means that hold-up issues (i.e. investment before the negotiation over a license) are endemic to the ICT industry. Furthermore, negotiations and litigation on infringement will typically be strategically timed to coincide with particularly large losses from product exclusion, i.e. to times at which hold-up power is particularly large.

C. Standard Essential Patents and Hold-Up

While hold-up issues are therefore pervasive in the ICT industries, they come to particularly sharp relief in the standard setting context. By agreeing on a standard (or when mandating it by regulation), no product for which the standard is essential for operation can be brought to market without depending on the use of the intellectual property. This means that without any commitment to pricing at the time of the adoption of the standard, a hold-up problem will also be endemic. The reason is that any user who has to invest in a product that uses the standard will need to negotiate ex-post with the relevant patent holders. By agreeing to the standard, investments have effectively been committed to if the firm wants to produce.

THAT HOLD-UP ISSUES (I.E. INVESTMENT BEFORE THE NEGOTIATION OVER A LICENSE) ARE ENDEMIC TO THE ICT INDUSTRY. FURTHERMORE, NEGOTIATIONS AND LITIGATION ON INFRINGEMENT WILL TYPICALLY BE STRATEGICALLY TIMED TO COINCIDE WITH PARTICULARLY LARGE LOSSES FROM PRODUCT EXCLUSION, I.E. TO TIMES AT WHICH HOLD-UP POWER IS PARTICULARLY LARGE.

In order to avoid this recognized hold-up problem standard setting, organizations (SSOs) have tried to establish rules that limit the ex-post monopoly power of patent holders owning patents that read on the standard: the commitment to (F)RAND rates (i.e. licensing rates that are Fair, Reasonable, and Non-Discriminatory). Unfortunately, this standard has proven to be highly ambiguous so that claims on FRAND rates on the same patent can vary by several orders of magnitude. Since there is also very large uncertainty about what the courts will determine as a FRAND rate there appear to be incentives to make extreme claims about FRAND. The uncertainty about what FRAND is and how the courts will determine it effectively undermines

the commitment and leads to a re-emergence of the hold-up problem: patent holders will exploit all of their bargaining power, which in circumstances in which courts are favorable and the litigation process favors patent holders can easily lead to considerable hold-up.

The discussions about SEPs have highlighted that hold-up can arise in two forms. One is through the award of excessive royalties by the courts. The second is through the process that leads to the determination of royalty rates. The latter is of crucial importance in real cases since the legal process will determine the outside options of the litigating parties and thus the bargaining power in any settlement process. For example Shapiro

has shown theoretically that unfettered injunctions will always lead to hold-up because they reduce the outside options of potential infringers in settlement negotiations relative to frictionless ex-ante negotiation process.⁸

To understand such biases against potential infringers, it is important to emphasize that patents in the ICT industry (and others) are probabilistic by nature. That the validity can never be satisfactorily determined ex-ante and that patent boundaries are vague leads to uncertainty about infringement on valid patents. A traditionalist view of patents might have assumed that all the bargaining power in such negotiations should go to the party who is holding the patent as long as it has been determined that the patent is not invalid. However, this forces the potential infringer to bargain as if the patent were valid for sure. As a result, the patent holder can extract far more in expectation than in a process in which validity and infringement could be determined instantaneously. This means the more the outcome of settlement bargaining can be tied to the actual final decision of a court or arbitrator, the more accurately the expected return reflects the actual social value of the patent. Rules that tilt the bargaining power toward the patent holder can significantly increase the hold-up problem and lead to a reduction in innovation in the market.

THE MORE THE OUTCOME OF SETTLEMENT BARGAINING CAN BE TIED TO THE ACTUAL FINAL DECISION OF A COURT OR ARBITRATOR, THE MORE ACCURATELY THE EXPECTED RETURN REFLECTS THE ACTUAL SOCIAL VALUE OF THE PATENT. RULES THAT TILT THE BARGAINING POWER TOWARD THE PATENT HOLDER CAN SIGNIFICANTLY INCREASE THE HOLD-UP PROBLEM AND LEAD TO A REDUCTION IN INNOVATION IN THE MARKET.

This insight explains the focus on injunctions in recent discussions on antitrust and other policy interventions. In particular, it underlies the concerns at the European Commission about the German Orange Book decision and its interpretation by the courts. Effectively a party that has been found to infringe can avoid an injunction by settlement, but this requires foregoing a later challenge of the validity of the patent. This will be true even if, as is often argued by its defenders, a preliminary assessment of likely validity is made by the court. Shapiro's analysis clearly implies that such a rule will lead to hold-up.

D. Asymmetries and the Hold-Up problem

Hold-up problems will not appear very severe when the different parties to the contracting are in a very similar position. Suppose that on average firms expect that that they will be as often (and severely) in a position to be held up by another firm as they are in a position to hold up the other firm. In terms of bargaining this evens the threat points. Indeed, an optimal solution is then to come to an ex-ante agreement to broadly cross-license the whole patent portfolio.

In the information technology sector such agreements have worked for a long time and allowed firms to innovate without having regard to potentially infringing patents of their main competitors. What has changed in the ICT industries is that the convergence and integration of different technologies—especially in mobile devices—has brought companies together with very different types of patent portfolios. The relative hold-up values of these portfolios are apparently much less clear across different previously non-integrated industries so that cross-licensing solutions become much harder.

In addition, cross-licensing solutions only tend to work between large firms with broad patent portfolios who are also practicing those patents. Whenever small companies need to negotiate with larger ones patent hold-up would re-emerge. Similarly the incentives of non-practicing entities, which do not produce products that might incorporate patents of other firms, do not have the possibility to trade on mutual hold-up opportunities. Hence, they will fully exploit the hold-up power that exists. To the extent that hold-up is an endemic problem in ICT industries, the tendency towards a greater role of non-practicing entities enforcing patent rights will lead to greater inefficiencies. In this sense clearer preventions of hold-up facilitating tendencies in the judicial process and hold-up preventing processes and rules for FRAND royalty setting (that remain in force even after a patent has been transferred) become even more important for preventing an increasing efficiency reducing impact of hold-up.

III. “SHOW ME THE HOLD-UP”: ARE WE FIGHTING THEORETICAL WINDMILLS?

As a result of the clearer conceptual focus on hold-up, many critics of policy intervention have withdrawn to the position that “there is no empirical evidence that there is hold-up.” In particular, it is often alleged that there is no evidence that hold-up has slowed innovation. This argument is akin to the old chestnut that there cannot be market power in the market because prices have been falling. The fallacy in all such arguments is that they ignore the fundamental concept of the counterfactual. Looking at a historical price path or a historical innovation path, one cannot determine whether price is higher or lower or innovation is higher or lower than the outcome in an efficiently operating market absent the market power effect.

The questioning of hold-up in the current discussion comes from a fundamental misunderstanding of valid economic evidence. The question is: how do we evaluate the observed outcomes in the market relative to an unobserved benchmark? In mergers, in which the competition authority is mainly interested in price effects, this is already difficult but not impossible to do empirically. Sometimes there are natural experiments in which a competitor is temporarily not in the market, is facing higher costs than the rival etc., which allows one to evaluate the degree of competitive pressure between rivals (from the variation in interaction).

But even in merger analysis some theoretical insight is necessary to interpret empirical results and guide the analysis on what one should look for. For example, the declining price fallacy arises from the belief that competition leads to falling prices. That is not the case. Productivity improvements lead to lower marginal costs, which will be passed on to consumers by firms. A monopolist will pass on cost reductions to some extent, because the optimal policy for rent extraction resolves a trade-off between a high price and high sales. But typically (although not always) competitive markets will have greater pass-through of cost reductions than a monopolized market. An increase in market power can thus be detected (after the fact) by changes in pass-through rates of cost reductions.

In most of antitrust analysis—especially when it comes to the impact of anticompetitive behavior on innovation rates (whether measured in R&D spending, or patent applications, etc.)—the counterfactual to the observed behavior is unobservable. The strongest evidence we have in these cases is a coherent theory of harm paired with evidence that the *assumptions* of the theory of harm apply to the industry and (at best) some indirect evidence that the anticompetitive effects appear to be a concern for the industry. This is the best one can do in terms of evidence in almost *any* antitrust case. Demanding a higher standard will make antitrust intervention

impossible. In fact, it probably would instead encourage more heavy handed regulatory intervention.

These limitations on feasible evidence do not mean that the standards are particularly low ones. It just means that the weight of theoretical considerations in the body of evidence has to be stronger. Indeed, the greatest challenge to antitrust authorities is often the formulation of a coherent theory of harm that satisfies basic economic logic and is consistent with incentives of the parties. Enforcement against selective distribution systems and other interventions on vertical relationships are examples among others where there are many antitrust interventions that are not based on *any* coherent theory of harm in the sense of analyzing the incentives of firms and therefore are not disciplined by a test of whether the central assumptions of the theory make sense in view of empirical evidence.

In the case of patent hold-up this is very different. We have excellent theoretical analyses of how hold up arises in patenting and that it is pervasive when certain conditions (like investments before license negotiation, certain features of the litigation process) are satisfied. Empirically, we have strong evidence that hold-up matters—and not just in standard essential patents. First, the RIM case has shown impressively how market costs of workarounds can create hold up as an ex-post phenomenon. We understand from empirical evidence that it is almost impossible to have patents that are not fairly uncertain in their validity or in their effective scope. Even extreme due diligence cannot avoid that a firm might reasonably conclude that it cannot possibly be infringing but the court disagrees. The great multitude of potentially relevant patents cannot all be reviewed if one wants to bring products to market in a timely way. This means that all the conditions for hold-up are generally present in the industry.

IN THE CASE OF PATENT HOLD-UP... WE HAVE EXCELLENT THEORETICAL ANALYSES OF HOW HOLD UP ARISES IN PATENTING AND THAT IT IS PERVASIVE WHEN CERTAIN CONDITIONS (LIKE INVESTMENTS BEFORE LICENSE NEGOTIATION, CERTAIN FEATURES OF THE LITIGATION PROCESS) ARE SATISFIED. EMPIRICALLY, WE HAVE STRONG EVIDENCE THAT HOLD-UP MATTERS—AND NOT JUST IN STANDARD ESSENTIAL PATENTS.

Similar conclusions apply for standard essential patents. If patents are essential for the implementation of the standard, they are unavoidable. Empirically, royalty setting occurs after the standard has been adopted and thus investments have been committed to. Such an institutional set up will necessarily lead to the hold-up problem. The fact that ex-ante (F)RAND constraints have been agreed to in standard setting organizations is evidence that members themselves recognize the hold-up problem. But the mere fact that there can be disagreements of several orders of magnitude on the “true” FRAND rate implies that the hold-up problem persists even with FRAND. To see this, note that parties would easily come to a settlement (or disagree less) if they had a clearer view of what courts would consider to be FRAND. The evidence thus shows that firms perceive courts not to put very predictable limits on FRAND rates inducing prolonged litigation. Again, based on the theory of hold-up and theories on why litigation persists, i.e. an analysis of incentives, it is obvious that the observed facts make it highly likely that the hold-up problem persists.

The patent cases in ICT industries that challenge the use of injunctions therefore satisfy the central requirements for a solid antitrust case. There exists a well specified theory of harm based on the analysis of incentives of

patent holders and potential patentees that implies hold-up and the associated harm. Second, there is considerable evidence that the general conditions under which one would expect such theories to apply are satisfied. There is thus coherence here between the theory, its assumptions, and the predictions it makes about market outcomes, which jointly generate a body of evidence that meets the typical standard of proof for antitrust cases; whoever thinks these do not should be honest enough to plead for a complete abandonment antitrust intervention against unilateral conduct. In fact, the standards of proof for regulatory intervention in markets tend to be far lower.

However, when one looks at the appropriate standard of proof one should not just address the question how strong the evidence for harm is. The appropriate standard should also depend on the potential social cost of intervention. If the social cost is potentially very high, one might want to have a higher standard of proof for the market failure that one is intending to address. In the antitrust cases on injunctive relief, there is strong evidence based on conceptual that the costs of suppressing interim injunctions should in most cases be small.⁹ But there is also the question on how important the patent system is in giving innovation incentives in the particular industry of concern. Views on this differ: There are some very strong views among academic economists that the empirical support for the effectiveness of the patent system is rather scant. Boldrin and Levin write: “The case against patents can be summarized briefly: there is no empirical evidence that they serve to increase innovation and productivity, unless the latter is identified with the number of patents awarded—which, as evidence shows, has no correlation with measured productivity.”

Just as in the broad claim on the empirical absence of evidence on hold-up, this statement has its problems as to the measurement against a well-specified counterfactual. But Boldrin and Levin may have a valid point that the value of the patent system appears to be considerably lower in ICT industries. For example, the explicit acknowledgement of many firms in the industry that they patent not in order to enforce their patents, but in order to threaten counter action should any one try to enforce theirs indicates that it is not the prevention of imitation that seems to determine patent behavior. Instead, it is a defensive use of patents to mitigate the risk from a firm’s product infringing on some patent of a competitor. Indeed, Google’s bid for Motorola was quite explicitly justified by the ability of Google to acquire a patent portfolio that could be used as a bargaining tool for IP cross-licensing. In fact, this is precisely the way one would expect patents to be used in an industry in which ex-ante contracting is difficult, hold-up is endemic, and defensive measures against hold-up become a central part of IP strategy.

IV. WHAT IS THE RIGHT POLICY INSTRUMENT TO ADDRESS HOLD-UP?

If one believes that on the basis of the current evidence some intervention to limit hold-up incentives has greater benefits on average than the potential expected costs of an intervention, there is still the question of how intervention should take place and what the best policy instrument is. Theoretically, it might even be the case that in the absence of potential regulatory intervention market incentives might be strong enough to create institutional solutions for the hold-up problem, but that the threat of intervention takes pressure of the actors in the market. In this section I will therefore explore the record on different potential solutions in order to better put the role of antitrust into perspective.

A. *Letting the Usual Suspects Resolve the Problem*

1. **IT IS A CONTRACTING PROBLEM: LET THE MARKET DO IT!**

While there is significant harm to consumers when hold-up problems are not avoided, this does not imply that intervention is needed. A naïve economist would think that the problem could easily be averted by writing better contracts at the beginning of the process. In other words, firms participating in a standard setting process should have an incentive to resolve hold-up problems through ex-ante contracting in the framework of standard setting organizations. Ex-ante contracting could then avoid any hold-up problem. Consequently, any observation of a lack of ex-ante contracting would simply reflect that the hold-up problem is not important enough to justify the relatively low transaction costs of negotiation.

However, a sound economic analysis is not just based on such theoretical considerations, but has to take into account the mounting empirical evidence that SSOs do not seem to be capable of agreeing on ex-ante terms even where extensive discussions take place. Secondly, it has to acknowledge that the evidence suggests that relatively costly and inefficient (temporary) exclusion from markets does occur and that there seem to be instances in which patent holders obtain very favorable deals for patents with dubious validity,

HOWEVER, A SOUND ECONOMIC ANALYSIS IS NOT JUST BASED ON SUCH THEORETICAL CONSIDERATIONS, BUT HAS TO TAKE INTO ACCOUNT THE MOUNTING EMPIRICAL EVIDENCE THAT SSOs DO NOT SEEM TO BE CAPABLE OF AGREEING ON EX-ANTE TERMS EVEN WHERE EXTENSIVE DISCUSSIONS TAKE PLACE.

suggesting a persistent hold-up effect. This means that the transactions cost of coming to ex-ante agreements of licensing appear to be much larger than the usual transaction costs of contracting. And if transaction costs are high enough, hold-up does indeed become inevitable and interventions to reduce it become ex-ante efficient.

The difficulty of obtaining agreements in standard setting organizations also appears to have good economic reasons: the negotiation of royalty rates before the adoption of a standard is problematic because it requires the revelation of research results before the standard is adopted. This revelation of information may reduce the ex-ante incentives for investment into intellectual property supporting a standard and make standard development harder. Quite typically applications for patents reading on a standard come very late in the process. Furthermore, in the standard setting process different players have systematically different incentives, which appears to be one factor that has prevented agreements on a sharper definition of FRAND and of a dispute process over FRAND in standard setting organizations like ETSI.

This does not mean that private contracting cannot or should not be part of a solution to the patent hold-up problems for standard essential patents, but the current evidence suggests that this will not occur without significant regulatory intervention that improves the incentives for coming to private agreements. The question then is not whether private contracting can lead to efficient outcomes, but what regulatory framework for SSOs creates the right incentives to come to agreements within SSOs that lead to both commitments to avoid the hold-up problem and incorporate the specialized knowledge of the industry to achieve efficient solutions.

2. LET THE COURTS DO IT!

Given the proven inability of private contracting solutions to resolve the hold-up problem it still appears *a priori* reasonable that the courts may be able to address the problem through adjudication of infringement, FRAND royalty fees, and determination of patent validity. In fact, the court system is the mechanism that is typically used to reduce the impact of incomplete contracting in by providing a way to resolve non-contracted for contingencies.

Indeed, even in the area of patent litigation, the US courts have gone very far in terms of limiting the elements of the litigation process that are most likely to increase the hold-up problem. Through the eBay judgment (and also through a recent opinion of Judge Posner in another matter), injunctions for patent infringements have been dramatically curtailed, essentially only applying when there is a real danger that the patent holder might not be able to recover damages in case of a finding of both infringement and validity. While these developments do not necessarily remove all issues concerning the determination of FRAND rates, a uniform application of the rules limiting injunctions would go a long way to limit the hold-up problem because they reduce the bargaining power of the patent holder only in the case of hold-up.

Unfortunately, the situation in Europe is much less promising. Different jurisdictions in Europe have adopted different rules. Some of these rules as those from the Orange Book judgment (in particular as applied by the Mannheim court) require giving up validity challenges in order to avoid injunctions and may require inordinate posting of bonds while the FRAND rate is determined. Such rules largely maintain the hold-up problem and lead to clear inefficiencies. The strong heterogeneity in court decisions and lack of clarity of rules therefore suggests that a reliance on the convergence of courts in different jurisdictions is unlikely, and a more direct regulatory intervention appears to be needed.

3. SOLVING THE PROBLEM THROUGH THE POLITICAL PROCESS

While an eventual resolution through the political process may be desirable, it is highly unrealistic that legislation would be adopted in the medium term. This would therefore not address the important market distortions we are observing at the moment due to the explosion in patent litigation in the ICT industries.

C. What Can Competition Policy Contribute?

In principle, the hold-up problem is a well-defined competition phenomenon involving an excessive amount of market power due to an increase in market power after a patent specific investment. Even where a patent should give monopoly power, it would be relative to the outcome of a complete market with ex-ante contracting. Hence, there is an excessive exercise of ex-post market power that is in play. The social cost comes both from the reductions in ex-ante incentives to invest in new products due to excessively high costs of patented inputs due to anticipated hold-up as well as the social costs generated from the disputes. These costs are exacerbated by temporary exclusion of products from the market, and its costs to consumers in form of less choice and higher prices due to reduced competition. As such, the hold-up problem can therefore be thought of as a classic competition problem to the extent that the market power exercised exceeds what you would get from the patent in a well-functioning world of complete contracts.

There are, in principle, two potential competition policy instruments that can address this issue under European antitrust rules that are based on a dominance standard. One can both construct arguments addressing the hold-up problem based on exploitative abuses (essentially excessive pricing) and on foreclosure abuses. I will here only discuss the excessive pricing abuses and will discuss the foreclosure arguments and the difficulty of implementing them further below when I deal with the monopolization standard used in the US.

An exploitative abuse can in principle be argued quite straightforwardly, since hold-up involves an excessive price relative to the appropriate ex-ante counterfactual of complete contracting. In addition, any actions that lead to excessive pricing have particularly high costs due to the reduction of innovation associated with hold-up in this industry. While this gives a stringent conceptual reason for applying exploitative abuse approaches to patent litigation cases in antitrust, there are some issues that might limit the appropriateness of the antitrust instrument.

First, the application of an abuse claim requires the existence of dominance. Such dominance claims have sometimes been ridiculed by lawyers in such cases. In a nutshell the question they raise is: How can you have dominance when the patent involved is only one of many? My impression is that this way of posing the question is a remnant of a traditionalist view of dominance that is not informed by economic reasoning. In modern antitrust we have come to see dominance as equivalent to strong market power. Patents licenses are complementary inputs for the products in this industry. With many necessary complements as input, each complementary input has monopoly power if there are no short run substitutes around. The dominance claim in these cases then becomes economically straightforward because the requirement of a standard turns the essential patent holder into an ex-post monopolist for any user of the standard (and on top of it one who has ex-ante promised not to exploit the monopoly position). This is true even if the patent is eventually shown to be invalid or inessential if an interim injunction can be used to influence ex-post royalty negotiations. Only when workarounds for a standard essential patent were possible at low cost for an infringing firm would a claim of dominance be problematic economically.

IN SOME CASES THERE ARE BEHAVIORS ASSOCIATED WITH EXCESSIVE PRICING THAT MAKE THE SCOPE FOR EXCESSIVE PRICING MORE SEVERE. IN SUCH CASES, INTERVENTION AGAINST SUCH BEHAVIOR CAN MOVE PRICES IN THE RIGHT DIRECTION WITHOUT THE NEED TO EXPLICITLY DETERMINE THE CORRECT PRICING BENCHMARK.¹⁰

Another issue with the excessive pricing approach is that hold-up needs to be measured against some benchmark. This benchmark problem is the typical problem of all excessive pricing (or exploitative abuse) cases. If the only solution comes down to explicit price regulation, this is often impossible to solve in a competition policy setting due to informational constraints. However, in some cases there are behaviors associated with excessive pricing that make the scope for excessive pricing more severe. In such cases, intervention against such behavior can move prices in the right direction without the need to explicitly determine the correct pricing benchmark.¹⁰

In principle, an excessive pricing standard could address both the procedural aspects that lead to greater hold-up as well as the determination of the FRAND rate itself. Given that the determination of a FRAND benchmark leads to the usual problems of price benchmarking in excessive pricing cases, limiting hold-up by addressing procedural issues like injunctions, dispute resolution, other ancillary rules in the standard setting context, or ways on how to commit to FRAND terms, appears to be a more promising way to move rates in the

direction of the theoretical ex-ante royalty rate in a complete contracts regime.¹¹ For example, threatening and/or using injunctions before validity has been resolved and rates have been set can be interpreted as a means of exploiting the judicial process to further enhance the ex-post bargaining power that allows the patent holder to achieve excessive royalties. A prohibition of asking for such injunctions (always under the appropriate caveats)

THERE MAY BE A BENEFIT OF GIVING GUIDANCE TO THE COURTS AND ESTABLISH A REASONABLE METHOD FOR DETERMINING A ROYALTY BASE, WHICH MAY MAKE REASONABLE FRAND RATE RANGES EASIER TO DETERMINE. IN PARTICULAR, IN THE ICT INDUSTRIES WHERE A MULTITUDE OF PATENTS RELEVANT TO ANY SINGLE DEVICE MIGHT MAKE IT HARD TO DETERMINE THE CONTRIBUTION OF ANY SPECIFIC PATENT TO THE VALUE OF THE PRODUCT, SIMPLE RULES BOUND THE DETERMINATION MIGHT AT LEAST REDUCE THE UNCERTAINTY OF COURT DECISIONS AND REDUCE LITIGATION COSTS THROUGH INCREASED CERTAINTY.

can therefore limit hold-up and thus push royalty rates in the right direction.

Of course, interventions that address the process by which royalty rates are determined instead of setting them directly may come with their own incentive problems. For example, the main obstacle to a straightforward prohibition of injunctions before determination of validity and level of the FRAND rate appears to be that a potential licensee may have an incentive to extend negotiations indefinitely in order avoid paying any royalties. This issue is typically discussed as the question whether the potentially infringing party is a “willing licensee.” The challenge for competition policy lies in finding criteria for willingness that make economic sense and allow mitigating the hold-up problem. Encouragingly, there seem to be some simple potential safe haven rules that have the structure of allowing an economically meaningful action by a potential licensee to avoid injunctions: For example, a binding commitment to submit FRAND determination to the courts

(or some other arbitration mechanism)—without losing the right to challenge validity. By establishing a safe haven rule of this type, competition policy can provide a framework, which can mitigate the hold-up problem and at the same time is well based in competition principles.¹²

It should be clear that a solution to the injunction issue does not necessarily eliminate all hold-up. But this does not necessarily speak against such an intervention. After all, the measure pushes settlements in the right direction and increases the probability of settlement by creating more legal certainty. It thus decreases the disincentive to innovation and reduces the impact of products being temporarily removed from the market.

But why not go further with an excessive pricing abuse claim and fix the problem of FRAND rate setting directly? After all this would address the hold-up issue head on and potentially resolve the thorny issue of what FRAND is? I believe that the problem here is a limitation of the effectiveness of competition policy instruments for the task at hand. Antitrust cases work well, where they resolve a particular issue in a particular case and give clear guidance to behavior in the future. The problem with determining FRAND rates is that the concept can be given clear meaning in terms of a theoretical counterfactual (i.e. rate that would have been obtained in a complete contracting world), but that there is no sufficiently simple algorithm to get to the right result in individual cases. An antitrust proceeding tends to be far too long and the informational problems so severe that determining FRAND seems outside the scope of antitrust authorities.

Nevertheless, it seems that also courts have difficulties to adjudicate on appropriate FRAND rates. Claims of parties about the appropriate FRAND rates can differ by several orders of magnitude, indicating that courts are sufficiently confused about the right order of magnitude for exaggerated to influence outcomes. Hold-up may thus be more severe because of a significant potential for excessive, but untestable, claims to succeed.

For this reason there may be a benefit of giving guidance to the courts and establish a reasonable method for determining a royalty base, which may make reasonable FRAND rate ranges easier to determine. In particular, in the ICT industries where a multitude of patents relevant to any single device might make it hard to determine the contribution of any specific patent to the value of the product, simple rules bound the determination might at least reduce the uncertainty of court decisions and reduce litigation costs through increased certainty. In particular, rules on the relevant royalty base for a patent could reduce the complexity of royalty determination and thus facilitate court decision making. Indeed, there may even be some good economic arguments that can be used to determine a reasonable royalty base.

However, there are significant practical problems for implementing a rule on royalty base with the tools of antitrust. Can the use of a particular royalty base by a patent holder ever be found an abuse of a dominant position? In principle, any base could be considered unproblematic if the royalty rate is appropriately adjusted. The difficulty for the courts is that without a common standard for comparing rates the appropriateness of the scaling is unclear. But giving guidance on such scaling is not necessarily a competition enforcement issue. It may therefore be difficult to use competition policy instruments beyond giving guidance on procedural aspects of FRAND in bounding the ranges for FRAND determination.

In principle, the standard setting organizations themselves would be in a much better position to determine rules of thumb that are appropriate for their own standard and use them as a more specific commitment than the vague notion of FRAND currently used. In the end, it is therefore likely that only multiple instruments can fully address the hold-up problems in patenting even when an excessive pricing instrument is available. Nevertheless, as we have seen, the abuse of dominance instrument can contribute something by limiting the competition distortions caused by the use of hold-up increasing procedural measures like injunctions.

Under a monopolization standard, the scope for intervention through competition policy instruments seems to be somewhat more limited. Excessive pricing cannot be captured by such a standard since the price setting behavior as such is not monopolization but just the use of a position of market power. It thus becomes central to construct arguments on the basis of an economically coherent foreclosure theory, which always is a difficult task.

A foreclosure argument requires proof that significant product innovations are likely to have been prevented by the possibility of hold-up or it would require actual foreclosure of competitors from markets to occur, for example, through an injunction. The problem is that non-investment in new products is very hard to prove. The foreclosure argument does also not exactly match the analysis of hold up as the central problem. If foreclosure is at stake, the demand of a royalty rate and subsequent request of an injunction are all geared to exclude a competitor from the market. This does not seem to exactly fit the problem that has generally been identified by those arguing for regulatory intervention. Similarly, there are many cases in which the patent holder and patentee are not in the same market (e.g. in the case of patent trolls). The typical foreclosure theories would then not apply. Since a monopolization standard makes essentially makes it necessary to argue foreclosure, the ability to address the patent hold-up problem will be more limited than under a dominance standard. It is

therefore not surprising that in US antitrust these issues have only been taken up by the FTC and under Section 5 of the FTC Act. It is a reflection of the fact that foreclosure theories of antitrust do not appear to be a good framework to address FRAND terms or the use of injunctions.

The above analysis has described a coherent defense of antitrust intervention through an excessive pricing instrument in the current patent disputes, but also shown that there are severe limitations to this instrument. In particular, it is hard to envision that competition policy could take a more direct role in FRAND determination.

V. COMPETITION ENFORCEMENT TOWARDS STANDARD ORGANIZATIONS

The limited ability of both the courts and the competition authorities to determine what FRAND rates are, points to the standard setting organizations themselves to address the issue in the context of the specific standard involved. Only in this way could a satisfactory commitment be created that reliably controls the hold-up problem. Unfortunately, standard setting organizations have a very poor record at coming to an agreement

IT IS POSSIBLE THAT A BETTER WAY OF INTERVENTION IS A REGULATORY FRAMEWORK FOR STANDARD SETTING ACTIVITIES.

on these issues—and not because they have not been discussed. The challenge for policy may therefore be to design policies that increase the incentives to come to meaningful agreements within the standard setting organizations themselves.

Could competition policy intervention towards standard setting organizations be such a policy route? Such a question is not all that outlandish given that SSOs consist of agreements between firms, many of which are competing in industries that are using the standards developed as inputs. Since certain types of rules facilitate hold-up and standards setting creates the monopoly power that comes with standard essential patents, the degree to which SSOs are beneficial to the competitive process rather than competition restricting agreements may depend on the specification of FRAND determination and dispute resolution rules in these organizations. It is possible that a better way of intervention is a regulatory framework for standard setting activities. But given the varied nature of standard setting bodies competition policy may have to have a complementary role. Indeed, recognizing this potential role may by itself further facilitate a resolution of the current patent disputes in the ICT industries.

VI. CONCLUSIONS

I have argued that, in terms of economic analysis, the hold-up problem at the heart of the patent abuse cases (especially in the standard essential patents) world can naturally be analyzed from a competition policy point of view. In an abuse of dominance framework, they fit most naturally into an exploitative practices category (i.e. a form of excessive pricing). Economic analysis makes a foreclosure argument that would be the only approach for a monopolization standard much less natural and far more difficult to prove. However, even with an exploitative abuse framework there are severe limitations on what competition policy can do in practice. Going beyond the banning of behavior that increases the hold-up problem, as the use of preliminary injunctions, appears to be almost impossible to achieve using competition policy instruments.

Competition policy intervention in Europe has so far confined itself to the limits suggested by this analysis.

It has taken a very limited role focusing exclusively on preliminary injunctions. Nevertheless, this may not fully address the hold-up problem. While the best way to resolve these issues is probably within the SSOs themselves, it appears that an appropriate regulatory framework that encourages agreement within these organizations is currently not in place.▲

1. University of Michigan, DICE, and CEPR
2. Recent publications by economic and legal scholars on the patent system speak of “Patent Failure” and “Patent Crisis”. (See J. Bessen and M.J. Meurer, “Patent Failure: How Judges, Bureaucrats, and Lawyers put innovators at risk”, Princeton University Press 2008; D.L. Burk and M.A. Lemley, “Patent Crisis and how the courts can solve it”, The University of Chicago Press, 2009).
3. M. Boldrin and D.K. Levine: “The Case Against Patents”, FEDERAL RESERVE BANK OF ST. LOUIS, Working Paper 2012-035A.
4. See Bessen and Meurer (2008), chapter 3, for an extensive discussion.
5. Indeed, Shapiro, Carl (2010), “Injunctions, Hold-Up, and Patent Royalties”, *American Law and Economics Review*, 12, pp. 280–318 shows that such hold-up will also occur with permanent injunctions
6. This does not mean that there is no distortion. Any contracting with a revenue or quantity based royalty will generate some degree of market power distortion. But this can be attributed to the way that the patent is meant to lead to a transfer of rent.
7. See Bessen and Meurer (2008).
8. Shapiro, Carl, 2010, “
9. See Ayres and Klemperer (1999), “Limiting Patentees Market Power Without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Relief”, *Michigan Law Review*, 96, pp. 985-1032. The main reason such suppression could be costly is when parties eventually cannot pay. However, there is no proposal in which this is not suggested as an exception to the general rule on injunctions envisaged.
10. For example, if a firm is a monopolist in country A but faces almost perfectly competitive conditions in country B. Assume that country B is also much larger. Then a requirement of non-discrimination will lead to approximately competitive pricing in country A (without moving the price in the much larger country much), leading to an improvement without determining the price level.
11. A discussion of different aspects of such procedural issues is given in Kühn, Scott Morton, and Shelanski (2013) (2013) 'Standard Setting Organisations can help solve the Standard Essential Patent licensing problem'. *CPI Antitrust Chronicle*, March 2013. Besides rules on injunctions these may include dispute resolution systems, rules on transfer of IP rights, and other aspects of the functioning of SSOs that could limit hold up problems.
12. Again there may be issues of preventing default of payment after a FRAND determination has been made that might require putting license payments in escrow before a final determination of the license fee liability. But these can in principle be incorporated into any safe haven rule of this type.

“Pay-for-Delay”: What Do We Disagree On?

BY PIERRE RÉGIBEAU¹

Antitrust concerns about “pay-for-delay” patent settlements are based on two theories of harms, one that stresses the need for courts to review the validity of patents and one that emphasizes the “probabilistic” nature of patent rights. The main weakness of the first theory of harm is that it fails to explain why some forms of patent settlements would be less desirable than others. The “probabilistic” theory of harm raises fundamental questions about the legal obligations of a patent-holder, the type of uncertainty that should be reflected in the probabilistic nature of the patents and whether the theory can be applied to anything but the simplest PFD settlements. This article also discusses the likely effect of a PFD ban on innovation and reviews both the European approach to recent and ongoing PFD cases and the recent *Actavis* decision of the US Supreme Court.

I. THEORIES OF HARM

“Pay-for-delay” or “value transfer” licensing settlements have been investigated by antitrust authorities in a number of jurisdictions, including the US, the EU and the UK. “Pay-for-delay” (PFD) refers to agreements reached between a pharmaceutical firm that produces a drug that is still protected by some patents and a (potential) generic entrant in settlement of litigation about the infringement and/or validity of the relevant patents. The key feature of a PFD agreement is that the generic agrees to enter only after a specified period and receives a positive transfer from the patent holder. The antitrust authorities’ objection to this type of settlement relies generally on two theories of harm.

According to the first theory, the fact that the patent-holder actually pays the generic challenger decreases the probability that patents will be effectively reviewed in court. As such review is an integral part of the patent system, this amounts to depriving society of the opportunity to “weed out” weak patents, thereby preserving unwarranted exclusion rights to the detriment of consumers.² The second theory of harm—commonly associated with the work of Carl Shapiro³—claims that any settlement in which the generic and the patent holder agree to allow independent generic entry prior to patent expiration but which involves a transfer from patent-holder to generic that exceeds the litigation costs expected by the patent-holder must involve a date of generic entry that is later than the “average” date that the patent-holder would—rightly or wrongly—expect to result from the continuation of litigation. This implies that the patent-holder essentially shares some of its expected monopoly rents to “delay” generic entry and that expected consumer surplus is lower than if litigation had been pursued to a final judgement. While this article tries to focus on issues that are common to both the US and the European “versions” of “pay-for-delay” cases, relatively more emphasis is given to the European approach. It is therefore important to remember that there is no equivalent to the Hatch-Waxman Act in Europe, so that keeping one generic out has no direct effect on keeping other generics out of the market.

II. THE MAIN POINTS OF DISAGREEMENT

I'll start with what I am not going to claim. I am not going to claim that any form of patent settlement should be acceptable under Antitrust Law. Clearly settlement terms that extend beyond the scope or lifetime of the patents deserve to be closely scrutinized. I would even agree that "value transfer" agreements should also be subject to some oversight as they could otherwise be used to protect extremely weak (or even sham) patents in a manner that would be hard to distinguish from a blatant market-sharing agreement. So, in my view at least, the main point of disagreement is not whether or not it is legitimate for competition authorities to be concerned about PFD agreements. Rather, the continuing disagreement between PFD "hawks" and "doves" stems from different views of the two theories of harm described above and of how these theories of harm can be applied to concrete cases. As I explain in more details below, the main question raised by the first theory of harm is why it should apply with special urgency to PFD deals. After all, *any* patent settlement effectively deprives society from the opportunity to invalidate patents that were granted in error. There are two main sources of disagreement about the second theory of harm. Firstly, as we will soon discuss, that theory is squarely based on a "probabilistic" view of patent protection. While undoubtedly appealing as a description of actual patent rights, the normative implications of the probabilistic view are far from being agreed on by all economists or legal scholars. Secondly, even if one were to subscribe to the probabilistic approach, it is not entirely clear what are the actual implications of this approach for patent settlements in general and PFD agreements in particular.

A. *First Theory of Harm: Are All Patent Settlements Objectionable?*

The review of patent applications by patent offices is necessarily imperfect, leaving a substantial portion of granted patents that are found invalid when further reviewed by the courts. This situation does not necessarily reflect poorly on the performance of patent offices. As the patents that are challenged in courts tend to be "those that matter," it is actually optimal to save the cost of a true in depth examination on the vast majority of applications. In this sense, then, ensuring an effective review of patents by the courts is important to the overall performance of our IPR systems. In spite of this, however, it is widely believed that litigation settlements have a useful role to play as they provide faster and cheaper alternatives to legal disputes. Either one believes that this principle also applies to the special case of patent litigation or one does not. If one does not, then any patent litigation settlement gets in the ways of socially useful judicial review. Our first theory of harm would then logically imply that *all* patent settlements, not just PFD settlements, should be prohibited.

If, on the other hand, one believes that settlements have a role to play in patent litigation, then the question is how one would distinguish between "good" and "bad" settlement. Following the logic of the theory of harm, bad settlements should be those that involve patents that are likely to be overturned by the court, i.e. "weaker" patents. The relevant question then is whether the presence of a transfer from the patent holder is a reliable indi-

...THE MAIN POINT OF DISAGREEMENT IS NOT WHETHER OR NOT IT IS LEGITIMATE FOR COMPETITION AUTHORITIES TO BE CONCERNED ABOUT PFD AGREEMENTS. RATHER, THE CONTINUING DISAGREEMENT BETWEEN PFD "HAWKS" AND "DOVES" STEMS FROM DIFFERENT VIEWS OF THE TWO THEORIES OF HARM DESCRIBED ABOVE AND OF HOW THESE THEORIES OF HARM CAN BE APPLIED TO CONCRETE CASES.

cator of the weakness of the patent involved in the litigation. The answer to that question is a qualified “no”. To see this, let us focus our attention on settlements that spell out a date of entry for the generic challenger as well as a possible transfer from the patent holder. Clearly the “overall package” offered to the generic must be more attractive if the patent is known to be weaker. This means that, if we compare two settlements with the same date of entry, one with a transfer and one without, one would generally believe that the settlement involving a transfer is associated with a weaker patent. On the other hand, we could not possibly draw any inference from the comparison of an agreement without payment and a given entry date and an agreement with payment that involves an earlier date of entry. So overall, when looking at a specific settlement, one simply cannot conclude that the presence of a payment implies that the patent at stake is weak. There no simple relationship between PFD and the strength of the underlying patent. Moreover, if one were to draw inferences from the combination of PFD and agreed upon entry date, one would face the following paradox: for a given size of transfer from the patent-holder to the generic entrant, a weak patent would lead to an *earlier* date of entry as the generic firm must be given a more attractive “package.” Clearly, a crackdown targeted at early entry agreements is not what competition authorities have in mind. Overall then, the first theory of harm does not seem to offer a sound basis for the singling out of PFD settlements.

B. Second Theory of Harm #1: Probabilistic Patents

It is important to clarify what economists mean when they refer to patents as “probabilistic” rights. There are essentially three “levels” of adherence to the probabilistic view.

Level 1: As a matter of positive analysis, the right to exclude granted by patents is without a doubt probabilistic as the patent-holder cannot be sure that the validity of the patent would be upheld if challenged and as, anyway, the precise coverage granted by the claims approved by the EPO remains quite uncertain until the construction of these claims has been further examined in court. Finally, even if there was no inherent uncertainty in the IP right itself, courts do make mistakes.

Level 2: As a matter of efficient design of a patent system, the probabilistic character of patent rights is actually desirable. As Ayres and Klemperer observe⁴, making patent rights “more probabilistic” is similar to reducing what economists refer to as the “breadth” of the patent. From the work of Gilbert and Shapiro⁵, we know that, under rather general conditions, a patent design that trades-off breadth against length makes it possible to ensure a given reward to the innovator at least social cost. So, at this second, normative, level, the probabilistic aspect of IPRs is useful, as long as IP owners are properly compensated by adjustments to, less distortionary, dimensions of the patent right (such as length).

Level 3: The probabilistic right is all that the patent-holder is entitled to. The patent-holder cannot therefore take any action that would eliminate the probabilistic aspects of the right if this action is to the detriment of consumers. Concretely then, a patent settlement will be seen as anticompetitive if it leads to a level of consumer surplus that is lower than the surplus that consumers would have expected as a result of continuing litigation. Assume, for example, that consumers would get a surplus of 50 if the patent-holder prevailed and did therefore continue as a monopoly supplier until the end of the litigated patent but that this surplus would increase to 100 if the generic entrant prevailed (say, if the patent was invalidated). If the ex ante probability of success of the patent-holder is p , then a settlement that leaves consumers with a surplus that is less than $50p+100(1-p)$ would be viewed as anticompetitive. This third, normative, view is what the second theory of harm relies on.⁶

While most economists have no problem with Levels 1 and 2, the same cannot be said about Level 3. Among the potential objections are:

Consistency. While they hold a valid patent, patent-owners who are not dominant are usually understood as facing no obligation to think about consumer welfare when acting within the scope of their patent. When entering a “normal” licensing agreement, for example, the terms of this agreement are properly set through bilateral negotiations without either party having to worry whether some alternative form of agreement would actually be better for consumers. Why then should such an obligation suddenly surface when an agreement—which might well involve licensing the technology—is reached as part of litigation against a potential generic entrant?

Practicality. A rule that says that a patent-holder can use and defend her patent while remaining within the scope of this patent as long as the patent is currently valid is easy for economic agents to understand and easy to enforce. A world where patent-holders would have to evaluate every substantial action regarding the use of their IPRs by assessing the fundamental uncertainty of their rights would appear to lack the clarity and predictability for which competition law should strive. If patent-holders really need to ensure that they always leave consumers with at least as much surplus as would result from actually “drawing” the “lottery ticket” that is a patent, how will patent-holders know what will or will not be deemed to be anticompetitive?

Sources of uncertainty. There are further degrees of “purity” even within the adherents to the “third level” described above regarding what should be seen as forming part of the “legitimate” probabilistic nature of patents, i.e. those probabilistic aspects that the patent holder should see as given and inalterable. Should the possibility of judicial error be included into the probabilistic nature of patents or should patent holders have the right to protect themselves against such error? What about injunction risk? In pharma, failure to obtain an injunction can be catastrophic for the patent-holder: generic entry will lead to lower prices and it is practically impossible to restore the pre-entry price level later on even if the patent-holder ultimately prevails.

WHEN ENTERING A “NORMAL” LICENSING AGREEMENT, FOR EXAMPLE, THE TERMS OF THIS AGREEMENT ARE PROPERLY SET THROUGH BILATERAL NEGOTIATIONS WITHOUT EITHER PARTY HAVING TO WORRY WHETHER SOME ALTERNATIVE FORM OF AGREEMENT WOULD ACTUALLY BE BETTER FOR CONSUMERS. WHY THEN SHOULD SUCH AN OBLIGATION SUDDENLY SURFACE WHEN AN AGREEMENT—WHICH MIGHT WELL INVOLVE LICENSING THE TECHNOLOGY—IS REACHED AS PART OF LITIGATION AGAINST A POTENTIAL GENERIC ENTRANT?

C. Second Theory of Harm # 2: What type of settlement should be prohibited under probabilistic patent approach?

Let us assume for the sake of discussion that we agree with the probabilistic patent benchmark: consumers should get at least what they would get if patent litigation was not settled. How do we know whether a given settlement satisfies such a criterion? This is where the work of Carl Shapiro becomes crucial as it is designed to provide us with a simple criterion, thereby addressing the “practicability” issue raised above. In a nutshell—and without getting into the myriad of possible variations on the model—when a settlement involves an agreed date

of independent generic entry, a transfer that exceeds the expected costs of litigation of the originator implies that the originator believes that the agreed upon date of entry is later than the expected date of entry if litigation

THE THEORIES OF HARM PRESENTED ABOVE TAKE AN EX POST VIEW: THE INNOVATION COVERED BY THE PATENT HAS ALREADY BEEN OBTAINED, SO THERE IS NO DISCUSSION OF HOW ANTITRUST ENFORCEMENT MIGHT AFFECT INCENTIVES TO INNOVATE. THIS IS AN IMPORTANT DRAWBACK. THE PATENT SYSTEM IS DESIGNED TO FOSTER INNOVATION AND ENSURES THE DIFFUSION OF KNOWLEDGE TO THE EVENTUAL BENEFIT OF CONSUMERS. AN ANALYSIS THAT IGNORES EFFECTS ON INNOVATION THEREFORE TAKES AN INCOMPLETE VIEW OF THE EFFECTS OF VALUE TRANSFER SETTLEMENTS ON WELFARE.

proceeded to the end. So, if one accepts the probabilistic patent benchmark, patent settlements involving such transfers can only be anti-competitive and should therefore be forbidden.

While useful, this criterion is not fool-proof even if one accepts the probabilistic patent benchmark. In particular it does not apply with such simplicity if the two parties have different exposure to risk or have different attitudes to risk. This later possibility should not be ruled out too easily, especially when the generic entrant is under severe financial constraints.⁷

Moreover, the Shapiro criterion simply does not apply to more complicated settlement agreements that do not simply involve an agreed date of independent generic entry with or without transfer. As just one example, there can be settlement agreements in which generic entry takes place immediately, but in which the generic must purchase from the patent holder at

an agreed transfer price or must pay an agreed royalty. Even if such agreements are accompanied by value transfers that exceed the expected future costs of litigation, it does not follow that such agreements necessarily reduce the expected welfare of the consumers of the affected drugs relative to expected consumer welfare if the parties had litigated. Since leaving consumers with at least the surplus that they could expect from continued litigation is what separates acceptable settlements from anticompetitive settlements under the probabilistic patent benchmark, the presence of a value transfer as part of such agreements simply cannot be seen as sufficient evidence that the agreement is anticompetitive.

In fact, it is easy to show that, for any date of generic entry that would be expected from the continuation of litigation, there is an immediate entry agreement with wholesale supply from the originator (or a royalty payment to the originator) that makes all parties—including consumers—better off. Moreover, in order to provide the generic with an incentive to enter this kind of welfare-enhancing agreement, a “reverse payment” from the patent holder to the generic will be required.⁸ It is therefore hard to see how the simple presence of a payment from the patent-holder to the generic could be used as evidence that any settlement involving immediate entry should be seen as anticompetitive.

D. A Policy Concern: Innovation

The theories of harm presented above take an ex post view: the innovation covered by the patent has already been obtained, so there is no discussion of how antitrust enforcement might affect incentives to innovate. This is an important drawback. The patent system is designed to foster innovation and ensures the diffusion of knowledge to the eventual benefit of consumers. An analysis that ignores effects on innovation therefore takes

an incomplete view of the effects of value transfer settlements on welfare.

There are two main issues here. Firstly, would banning value-transfer settlements actually hurt the profits of patent-holders in the pharmaceutical sector? This is not clear *a priori*. On the one hand, value transfer settlements enable the patent-holder to settle litigation at a lesser cost, so removing this possibility would hurt. On the other hand, generics also benefit from value transfers so they might be less willing to challenge the patent in the first place if the practice was removed. That effect would be beneficial to patent-holders. If the net effect is of banning value transfers is favourable to patent-holders, then we are effectively in a situation where patent-holders' own ex post rational use of value transfers militate against their ex ante interest. Is this really ground for pursuing them under antitrust law? If on the other hand, the net effect of banning value transfers is to decrease the patent holder's expected profits, then one must consider the feedback effect on innovation.

Let us first dispose of a red-herring. Authors on both sides of the debate have invoked the fact that "the patent system has been designed to balance a variety of effects optimally," so one should not unduly tinker with it through competition law. On the "anti-transfer" side, the argument is that the patent system is optimal in an environment where value-transfer settlements are not allowed, so there is no need to compensate innovators if value transfer settlements are banned. On the other side, the (implicit) view is that, since opposition to value-transfer settlements is recent, one must assume that the patent system balanced effects under the assumption that all kind of settlements within the scope of the patent would be allowed. In this view, banning value transfer settlements would therefore have potentially serious effect on the balance of the IP system in pharmaceuticals. Such debate over original intent is useless.

If one wants to study rigorously the effect of value-transfer settlements within the patent system, one must follow the usual approach used in the patent design literature evoked above: take the level of reward to innovators as given and determine whether allowing for value transfer settlements makes it more or less costly to consumers to provide this level of reward to innovators. If one conducts that analysis, one actually finds that a value transfer settlements make it possible to provide a given reward to consumers at a *lesser* cost in terms of ex post consumer welfare.⁹ In other words, even within a patent system designed on the basis of a probabilistic view of patents, there would still be room for allowing for settlements that involve payments from the patent-holder to the generic.

The second issue is whether banning value-transfer settlements would actually address the two theories of harm described at the beginning of this note. In particular, it is far from clear that a prohibition on value-transfers would actually lead to a more efficient "weeding out" of bad patents. Just as the impact of a prohibition on innovation incentives was ambiguous, one cannot conclude that it would lead to more challenges working their way to a final litigation outcome: a ban on value transfers might increase the proportion of generic challenges making it all the way through litigation but, since it decreases the expected pay-offs of the challenging generic, it could also lead to fewer challenges in the first place.¹⁰

III. THE EUROPEAN APPROACH

DG Comp has been pursuing a small number of cases. If we look across these cases, we notice that the Commission is relying on both Article 101 and Article 102, sometimes within the same case. Under Article 101, the Commission sees value-transfer settlements as "per object" infringements. The Commission's approach

under Article 102 is less clear. Still, while the Commission does present arguments relating to the therapeutic substitutability of the medicine at stake with other medicines within the same class, there appears to be a new emphasis on a definition of dominance that is arguably “tailored to the alleged abuse.” Under this approach, the very fact that the patent-holder had the power to exclude generic competition through value-transfer “bribes” and that it had the incentives to do so—since prices tend to fall abruptly when generics enter—suffices to establish dominance, irrespective of the extent of therapeutic substitutability within class, the intensity of non-price competition or the overall profit margins realized on the protected drug.

The Commission has reached a decision in the Lundbeck case, imposing rather large fines on both Lundbeck and the generic companies involved in the PFD agreements. It is worth noticing that the decision relies exclusively on Article 101, even though Citalopram’s shares in its therapeutic class were at least as high as the shares of some other medications for which Article 102 is also used. The main reason for this appears to be the urgency of closing the case to avoid running into the status of limitation.

A. Collateral Damage: Difficulties in Applying Articles 101 and 102

As mentioned above, the European Commission has at times relied on both Article 101 and Article 102 to pursue value transfer settlements. Each approach presents its own difficulties.

When using Article 101, the main difficulty comes from determining whether generic companies can be seen as “potential competitors” as long as the patent at stake is valid. Traditionally, firms that are barred from entering a market because of the presence of a valid patent have not been seen as potential competitors in this market. Since a patent is presumed valid until it is voided by a competent authority, a generic challenger which settles with the patent-owner cannot then be seen as a potential entrant since, if there is indeed infringement, at the time of the agreement entry could only occur in violation of the patent. In a sense, then, the traditional view of a patent as being either “on” or “out” is mirrored by a dichotomous assessment of potential competition as “on” if entry does not violate a “on” patent and “off” if it does. This suggests that the pursuit of PFD cases under Article 101 requires a redefinition of the notion of potential entry to fit the probabilistic theory of harm: if patents are thought to be probabilistic, then it would also make sense to consider potential entry as a probabilistic concept. In that view, a generic entrant would still be seen as a potential competitor if there is a sufficiently high probability that it would actually prevail in litigation and therefore be able to enter the market. An interesting implication of this view would be that Article 101 could only be applied if there was sufficient evidence that the patent at stake is weak. However, the European Commission has carefully avoided any reference to the strength of the patents involved in PFD deals and has certainly not presented any evidence suggesting that those patents were weak. In my opinion, this is inconsistent with the need to redefine the notion of potential entrant in a probabilistic manner that fits the probabilistic nature of the Commission’s theory of harm.

The use of Article 102 raises two main issues. The first one is the traditional unease that some observers feel when abuse of dominance is used to get at an *agreement* between willing parties. The second relates to the manner in which dominance is established. We do not need to discuss here the general issue of how one assesses market power and dominance in “high sunk cost” industries such as pharmaceuticals, since this is not specific to value transfer settlements. However, the Commission’s approach to dominance seems, as mentioned above, to be tailored to the specific abuse that it pursues. In a nutshell, the Commission considers that the fact that generic production of a given medication leads to a collapse in the price of the drug, while generic entry into

drugs that are good therapeutic substitutes does not, is evidence that, for the type of abuse considered, each drug is a market onto itself, regardless of how many close therapeutic substitutes are available. This seems to boil down to saying that any patent-holder who is the sole supplier of a drug that sells for a price that is substantially higher than its variable cost of production will be found to be dominant. In practice, that implies that, in the context of alleged abuses regarding generic entry, the vast majority of existing drugs confer dominance on the relevant patent-holder.

The Commission's approach raises two main issues. Firstly, should dominance be assessed solely in terms of price behavior? Given that, in most health systems, doctors and patients have little incentive to consider the price of the medicines that they prescribe or use, the fact that a decrease in the price of a drug has little effect on the price or sales of another drug that is a close therapeutic substitute is hardly surprising. However, under European law, dominance is defined broadly as the ability to behave to a substantial extent independently from other firms and consumers. This definition seems to imply that all forms of competition should matter. In particular there is intense rivalry between therapeutic substitutes in terms of "share of voice" (i.e. medical profession advocacy), experimental studies and research. Disregarding these dimensions of competition to narrowly focus on a price rivalry that is inhibited by the rules of the health system seems hard to justify.

...FINDING A DRUG DOMINANT WHENEVER GENERIC ENTRY WOULD LEAD TO A SUBSTANTIAL DECREASE IN PRICE AMOUNTS TO EVALUATING THE MARKET POWER OF THE DRUG COMPARED TO A COMPETITIVE BENCHMARK WHERE PRICES ARE EQUAL TO MARGINAL (OR AT LEAST VARIABLE) COSTS. THIS MAKES NO SENSE IN AN INDUSTRY WITH HIGH SUNK COSTS. IN SUCH INDUSTRIES, A MUCH MORE NATURAL COMPETITIVE BENCHMARK IS THE PRICE AT WHICH THE DRUG MANUFACTURER BREAKS EVEN OVER THE LIFETIME OF THE PRODUCT.

Secondly, finding a drug dominant whenever generic entry would lead to a substantial decrease in price amounts to evaluating the market power of the drug compared to a competitive benchmark where prices are equal to marginal (or at least variable) costs. This makes no sense in an industry with high sunk costs. In such industries, a much more natural competitive benchmark is the price at which the drug manufacturer breaks even over the lifetime of the product. Such a benchmark at least ensures that a firm producing a drug that does not even recover its initial R&D expenditure is not found "dominant" and hence endowed with the special responsibility that comes with dominance. Establishing dominance based on ex post marginal cost pricing makes a mockery of the whole intellectual property bargain whereby patent protection aims at rewarding inventors and hence at insuring that, on average at least, they can at a minimum recover their initial investment.

B. A Broader Policy View

It is also interesting to ask what the likely impact of banning value-transfer settlements might be. We have already discussed what the potential effect on innovation might be. We now turn to the likely effect on settlements and a potential effect on the behaviour of generics companies.

After completing its review of the pharmaceutical sector in 2008, in which it indicated that it saw value transfer settlements as potentially problematic, the Commission decided to keep track of pharmaceutical settlements, classifying them in three categories (see table). The Commission concluded, with some satisfaction,

that the proportion of settlements that imposed no restrictions on generic entry had increased. Moreover the proportion of cases that limited generic entry but without transfer payment increased compared to cases where entry was restricted and transfer payments were made.

The Commission concluded that its negative stance on value transfer settlements had not made it more difficult for firms to settle and had been effective in reducing the occurrence of the objectionable kind of settlement. The first point is of course not correct: the fact that there is still a large number of settlements tell

...EVERY SINGLE CASE IN THE CATEGORY INVOLVES A SETTLEMENT WHERE THE GENERIC ENTRY WAS DELAYED UNTIL THE END OF THE PERIOD OF PATENT PROTECTION. IN OTHER WORDS, OVER FOUR YEARS AND MORE THAN 400 SETTLEMENTS, THERE WAS NOT A SINGLE EXAMPLE OF THE TYPE OF SETTLEMENT WHERE GENERIC ENTRY IS ALLOWED AT A DATE THAT IS SUPPOSED TO REFLECT THE PARTIES' APPRAISAL OF THEIR RESPECTIVE CHANCES AT TRIAL.

us nothing, without any information on the population of actual and potential litigation cases that these numbers refer to. As for the second conclusion, there is a bit of a sleight of hand. First, cases of settlements without limit on generic entry are almost all cases where the patent-holder had already essentially lost the case because of adverse preliminary rulings. Second—and most interestingly—every single case in the category involves a settlement where the generic entry was delayed *until the end of the period of patent protection*. In other words, over four years and more than 400 settlements, there was not a single example of the type of settlement where generic entry is allowed at a date that is supposed to reflect the parties' appraisal of their respective chances at trial.

Table 1: Type of Pharmaceutical Patent Litigation Settlement Before and After the European Commission's Sectoral Inquiry

	Total settlements	Settlements, by type		
		No limit on generic entry (A type)	Generic entry restricted, but no value transfer from originator to generic (B.I type)	Generic entry restricted, value transfer from originator to generic (B.II type)
Jan 2000 to June 2008 (Inquiry)	207	104 (50%)	54 (26%)	46 (22%)
Jul 2008 to Dec 2009	93	53 (57%)	31* (33%)	9 (10%)
Jan 2010 to Dec 2010	89	54 (61%)	32* (36%)	3 (3%)
Jan 2011 to Dec 2011	120	84 (70%)	23* (19%)	13 (11%)

This raises two questions. Firstly, where are the welfare benefits from the Commission's expressed negative view of value transfer settlements then? Have private parties just gotten cleverer without any benefit for consumers? Secondly, does this raise a question about how appropriate the counterfactual used in the "Shapiro" theory of harm? The counterfactual where parties agree on an entry date that reflects the strength of the patent is a fine theoretical benchmark, but is it a useful policy benchmark if it is never observed?

IV. THE SUPREME COURT *ACTAVIS* DECISION

One of the very few things that are clear in the recent Supreme Court Decision is that the Court could not follow an extreme "scope of the patent" approach that would have made impossible to guard against disguised "market sharing agreement. However, the Court failed to identify clear reasons why PFD might be considered to be anticompetitive. For example, while the Court recognises that patents need to be further tested in Court, it fails to explain why this might imply that some patent settlements are lawful while others are not. It seems also clear that the Supreme Court did not embrace the extreme "probabilistic patent" approach. While the Court clearly sees patents as probabilistic in our "Level 1" sense and points out that reverse payments are puzzling, there are no references to the probabilistic benchmark (our "Level 3") according to which consumers should get at least the level of welfare that they would expect from continued litigation. Overall, then, the Court appears to have opted—not surprisingly—for a very traditional approach: the enforcement of IPRs, including settlements, is not a matter for patent law only (even if it allows for antitrust considerations) but is potentially fair game for antitrust authorities. It does not mean that enforcement will be simple however as the Court acknowledges that agreements that include reverse payments cannot be seen as "presumptively unlawful." Interestingly, the Court seems to recognize the importance of patent strength in establishing whether or not an agreement is lawful under a rule of reason, going as far as pointing out that the relationship between the size of the payment and the implied strength of the patent would be one of several pieces of information that would make a detailed analysis of patent validity unnecessary.

THE COMMON DENOMINATOR OF THE COURT'S CONCERNS IS THAT THEY ARISE MOSTLY WHEN THERE ARE REASONS TO BELIEVE THAT THE PATENTS UNDER LITIGATION WOULD BE LIKELY TO BE OVERTURNED IF LITIGATION PROCEEDED TO THE DIRE END. IN THAT SENSE, THE *ACTAVIS* DECISION SEEMS TO BE AT ODDS WITH THE APPROACH OF REGULATORS LIKE THE EUROPEAN COMMISSION WHO HAVE GONE TO GREAT LENGTH TO KEEP THE NOTION OF "PATENT STRENGTH" OUT OF THEIR LINE OF ARGUMENT.

Clearly, then, the Court condoned neither an impervious "scope of the patent" approach nor an extreme "probabilistic" view. In this sense, the decision is compatible with our previous discussion as neither of these views is a realistic basis for effective policy. The first one ignores the real concerns that patent settlement agreements can support market-sharing deals and the second one is both too extreme in its logical implications for other aspects of licensing behaviour and would be basically impossible to implement in all but the simplest cases. One could however interpret the Court's current position in light of the two theories of harm that we have discussed. In terms of the first theory, the Court sees the need to ensure the review of patents, especially when these are likely to be weak. In terms of the second theory of harm, the Court seems to be most concerned

by the fact that PFD settlements could be used to preserve the unjustified monopoly rents of a weak patent. The common denominator of the Court's concerns is that they arise mostly when there are reasons to believe that the patents under litigation would be likely to be overturned if litigation proceeded to the dire end. In that sense, the *Actavis* decision seems to be at odds with the approach of regulators like the European Commission who have gone to great length to keep the notion of "patent strength" out of their line of argument.

Unfortunately, the *Actavis* decision does not tell us much more than that. The decision is particularly obscure in terms of burden of proof. While most of the language suggests that the burden of proof is essentially on the FTC, the Court also seems to leave the door open to a claim that a very large reverse payment would itself be presumptive evidence that would then need to be refuted by the defendant. Furthermore, if one were to actually read the decision as establishing a "rule of reason" approach, there is very little guidance as to the type of evidence that "reason" should look at. The decision is particularly ambiguous as to the role of patent "strength." Is the demonstration that the patent could reasonably have been seen as strong at the time of the agreement always a legitimate defence or is it trumped anyway if the transfer from the patent-holder to the generic is judged to be "unreasonably" large anyway?

Overall then, even though the reasons for disagreement between value transfer "hawks" and value transfer "doves" are by now fairly clear, I fear that the *Actavis* decision has done little to bring about a quick resolution of the PFD debate.

V. CONCLUSION

Competition authorities have relied on two main theories of harm to pursue PFD settlements. The first theory states that such settlements unduly prevent the patents at stake from being properly re-evaluated by a Court. The main weakness of this theory of harm is that it fails to explain why PFD settlements should be seen as less desirable as any other type of patent settlement. The second theory of harm relies on the view of patents as probabilistic property rights. This theory of harm has two anchors. The first one is the claim that the holder of a probabilistic right should ensure that consumers enjoy a level of welfare that is at least as high as the level that they would expect from the completion of patent litigation. The second anchor is the analytical result proposed by Carl Shapiro which shows that, under some conditions, the mere presence of a transfer from licensor to licensee that exceeds the expected litigation costs of the licensor is sufficient to establish that consumers lose from the settlement. While this theory of harm is worth taking seriously, it has a number of weaknesses, including inconsistencies between the probabilistic view and traditional antitrust treatment of licensing, the fact that it cannot be applied to more complex agreements where payments are accompanied by immediate generic entry and the identification of the sources of uncertainty that are properly reflected in the "probabilistic" nature of the patent rights. The overall effect of policies banning PFD settlements on innovation is also a concern.

Turning to the ongoing investigation of PFD agreements in Europe, I briefly discuss three sources of controversy. The first issue is what the proper definition of a "potential entrant" should become when one considers the patent rights themselves to be probabilistic. I argue that the logical approach would be to adopt a probabilistic definition of potential entry itself but that this also implies that only settlement of litigation relating to patents thought to be weak should be a concern. The second issue is the approach currently taken to determine dominance in PFD cases. I argue that this approach not only relies on the wrong competitive benchmark but it simply ignores the strong competitive constraints that therapeutic substitutes exercise on each other through

non-price channels. Finally, tracking the evolution of settlements since the European Commission’s review of the pharmaceutical sector—where doubts about the legality of PFD settlements were first expressed—shows that the kind of settlement where firms agree on a date of generic entry without side payments actually do not arise. Since this type of settlement is the benchmark compared to which PDF agreements are thought to be abusive, this raises questions about the very foundations of the Commission’s theory of harm.

Finally we argue that the recent *Actavis* decision does not support either a pure “scope if the patent” approach or a pure “probabilistic patent” approach. Rather it seems to attempt to strike for a middle ground where the strength of the patents at stake would be an important element of the competitive approval of PFD settlements.▲

1. CRA and Imperial College. This paper reflects the opinions of the authors, not those of CRA. The authors has been representing patent-owners in European litigation.
2. Contrary to what is sometimes stated, this theory of harm is not contingent on the “quality” of the patent system: even in the best possible system, it is optimal to limit the resources devoted to the review of patent applications and let further administrative or judicial processes refine the review of patents that prove to be sufficiently controversial and/or important to be the object of post-grant challenges. Of course, the expected *level* of harm depends on how much the system relies on such ex post mechanisms and hence on the initial quality of patent review.
3. C. Shapiro, 2003, “Antitrust Limits to Patent Settlements”, *Rand Journal of Economics*, 34:2, pp. 391 – 411.
4. Ayres, I. and P. Klemperer, 1999, “Limiting Patentee’s Market Power Without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Remedies”, *Michigan Law Review*, 97:4, pp. 985 – 1033.
5. Gilbert, R.J. and C. Shapiro, 1990, “Optimal Patent Length and Breadth”, *Rand Journal of Economics*, 21:1, pp. 106 – 112.
6. One can take the view that the Level 3 interpretation of probabilistic patents is indeed a consequence of level 2. It is *because* probabilistic patents might be part of an optimal patent system that it is important to hold companies to that standard. This of course assumes that the current patent system was indeed build in accordance with what economic theory tells us about the role of such a probabilistic dimension.
7. The possibility of bankruptcy tends to make firms behave in a more “risk-loving manner”.
8. The analysis supporting this statement can be found in the working paper version of this article, P. Régibeau, “Pay for Delay: What Do We Disagree On?”, SSRN Working Paper #2368220
9. The formal analysis underlying this point can be found in the working paper version of this article. See footnote 7.
10. A formal analysis of this point is available in the working paper version of this article. See Footnote 7.

Pay-for-Delay

BY FIONA SCOTT MORTON¹

This article lays out the economics of competition between branded and generic pharmaceuticals and its welfare consequences. I explain the logic behind so-called “pay-for-delay” or “reverse payments” in the context of the current IP environment where weak (probabilistic) patents are frequently granted by the PTO. The article goes on to relate the Supreme Court decision in *Activis* to these concepts. I argue that the “scope of a patent” is closely related to its probability of being valid. The Supreme Court dissenting opinion states that IP owners should be allowed to operate within the scope of the patent. For a very weak patent, that might be a very limited scope and bring the dissent into agreement with the majority opinion that a weak patent owner should not be allowed to create market power where the patent did not grant it. However, the dissenting opinion closes with a rejection of using the concept of probabilistic patents in legal analysis.

I. ROLE OF THE PHARMACEUTICAL BRAND

The brand receives patent protection in order to incentivize innovation. As is well known, and does not bear covering in detail here, society developed the patent system in order to give innovators property rights that create incentives for innovation. This initial period of patent protection is likely to be important for encouraging innovation in pharmaceutical treatments. There is considerable economics literature that asks whether the patent system is a net drain or stimulant to innovation, but it is widely believed that the pharmaceutical industry is one of the most positive cases.

A successful brand has invested in clinical trials and other activities to obtain regulatory approval. In addition to basic research and early clinical trials, the regulator in the US or Europe requires extensive clinical trials to determine safety and efficacy. These often involve thousands of patients and take many years. The financial investment in bringing a new pharmaceutical product to market is therefore considerable (recent US estimates approach \$1 billion).

At the end of the approval process the innovator typically has a patent on the original molecule as well as additional patents on other aspects of the brand, such as a pill shape or extended release formulation. The innovator also may have market exclusivity, which in the US is a guaranteed minimum number of years before generics may enter. Lastly, the innovator may have data exclusivity on the results of its clinical trials for a period of time. Data exclusivity means that no other firm can use the trial data showing the effectiveness of the drug during the time of exclusivity.

II. SOCIETY WANTS COMPETITION FOR THE BRAND ONCE THE PATENT EXPIRES

The idea of a patent is to reward the innovator with a limited-time exclusivity on its invention after which the innovation is available without cost to society. A generic entrant that neither has to pay the fixed cost of inventing the drug, nor the fixed cost of determining that it works in real people, faces only the low marginal cost

of manufacture. Such a generic can cover its costs by selling the drug for a price close to this marginal cost, and in that way benefit society by greatly expanding access to the product, and, in many modern economies, reducing the tax burden needed to finance the public healthcare system. Generic entry into pharmaceutical markets where patents have expired is therefore desirable (assuming optimal patent life).

In the US, legislation explicitly designed to achieve this goal, the Hatch Waxman Act (HW), was passed in 1984. It established an exclusivity period for innovators, thereby guaranteeing the brand a minimum period when it could sell without facing competition no matter how slow the FDA approval process. It also reduced the fixed cost of entry for generics by requiring only that they demonstrate a product is bioequivalent to (almost the same as) the brand, rather than running long safety and efficacy trials. This shorter procedure was called an 'Abbreviated' New Drug Application, or ANDA. Prior to HW, if a generic challenged a brand's patent and the patent was found invalid, the market would immediately be open to all generic entrants. Because competition is fierce in the generic industry, profits would be low and provide no return on the cost of the initial litigation for the challenger.

HW fixed this problem by providing that the first generic to defeat a brand patent be given six months of generic exclusivity. These generic entrants are known as "Paragraph IV" entrants because upon entry, they certify to the FDA that they are infringing a brand's patent but assert that the patent is not valid or not infringed. The first paragraph IV filer that succeeds in proving its case gains the right to be the only generic on the market for six months (the FDA will not approve others), during which time it competes only with the brand. These six months are lucrative, and therefore provide the financial incentive to litigate a weak brand patent. After six months, other generics may enter and society gets the benefit of the fact that the patent has been shown to be invalid. The critical feature of the legislation is that should the initial litigant not enter, the six month period does not begin and no other generic can enter, either. Thus without the first generic entrant, the brand faces no generic competition at all. Moreover, only one Paragraph IV exclusivity period is granted under the law; no subsequent ANDA-filer may earn it, and therefore no subsequent generic entrant has a strong financial incentive to litigate. Settling with the first Paragraph IV generic entry therefore strongly reduces the likelihood of subsequent generic entry until all the brand's patents have expired.

THE CRITICAL FEATURE OF THE LEGISLATION IS THAT SHOULD THE INITIAL LITIGANT NOT ENTER, THE SIX MONTH PERIOD DOES NOT BEGIN AND NO OTHER GENERIC CAN ENTER, EITHER. THUS WITHOUT THE FIRST GENERIC ENTRANT, THE BRAND FACES NO GENERIC COMPETITION AT ALL. MOREOVER, ONLY ONE PARAGRAPH IV EXCLUSIVITY PERIOD IS GRANTED UNDER THE LAW; NO SUBSEQUENT ANDA-FILER MAY EARN IT, AND THEREFORE NO SUBSEQUENT GENERIC ENTRANT HAS A STRONG FINANCIAL INCENTIVE TO LITIGATE. SETTling WITH THE FIRST PARAGRAPH IV GENERIC ENTRY THEREFORE STRONGLY REDUCES THE LIKELIHOOD OF SUBSEQUENT GENERIC ENTRY UNTIL ALL THE BRAND'S PATENTS HAVE EXPIRED.

After HW took effect the number of generic entrants in US pharmaceutical markets increased, generic entry immediately after patent expiration became frequent (often measured in days), and generic penetration

grew steadily to the point where generics now routinely capture more than 90 percent of all prescriptions for a given molecule. Research shows that generic prices decline with the number of entrants, typically beginning at about 80 percent of the branded price and falling to as little as 10 percent of the brand's price in large markets with a dozen or more entrants. The statistics in Europe are a little less favorable (the generic often being the local firm, and pharmacies often successfully lobbying for high margins). However, generic entry after patent expiration in Europe is similarly beneficial for consumers and healthcare budgets.

III. TEMPTATION

The cost structure of the pharmaceutical industry is very important in shaping legal strategies. As noted above, the brand's price is typically five-to-ten times the generic price. The brand's gross margin is therefore very high, often around 90 percent. Generic entry takes sales away from the brand, often in a dramatic and significant manner. The generic, due to competing with a homogeneous product in an industry with low barriers to entry, does not earn large economic profits. Indeed, price is often close to marginal cost and therefore the generic saves the consumer much more in consumer surplus than the generic collects in profits. To fix ideas, imagine the brand's price is 100, the generic price is 20, and manufacturing cost is 10. On each unit sold the generic earns only 10 while saving the consumer 80 and costing the brand 90.

Even the first generic into the market does not earn the high economic profits typical of a brand, and thus it is easy for the brand to make the generic an attractive offer while leaving ample profit for itself. The brand

TO FIX IDEAS, IMAGINE THE BRAND'S PRICE IS 100, THE GENERIC PRICE IS 20, AND MANUFACTURING COST IS 10. ON EACH UNIT SOLD THE GENERIC EARNS ONLY 10 WHILE SAVING THE CONSUMER 80 AND COSTING THE BRAND 90.

offers X percent of its remaining brand profit to the successful Paragraph IV generic entrant—if it agrees not to enter the market. Since other generics may not enter until the initial paragraph IV entrant has enjoyed its six months of exclusivity, and have little incentive since they will not receive any exclusivity protection, entry is effectively blocked. This leaves $100-X$ percent of

profits for the brand, rather than almost zero, which is what it would receive should generics enter. The parties in this way can divide up the monopoly profits in the market. The parties prefer to settle patent litigation in this manner because it preserves the brand's monopoly profit for division between the two firms; the missing element, of course, is that consumer surplus is not considered in the arrangement. The profit accruing to both parties would be shared with consumers to a significant degree if the usual entry process took place.

The key to the strategy is found in the HW regulation: FDA approval of subsequent generics requires the Paragraph IV generic entrant to sell its product for six months, and therefore the first generic entrant can block subsequent entry by failing to enter the market. In a market without regulated entry, subsequent entrants would face the same costs and rewards as the first 'entrant' -- who chose to stay out. In the classic pay-for-delay scenario, the brand and the generic settle their patent litigation by agreeing that the generic will stay out of the market for a certain number of years while the brand pays the generic a lump sum. If the generic had infringed a valid patent, and owed damages, one might expect settlement to involve a payment in the reverse direction, from the generic to the brand; this is why these settlements are called "reverse payments." The delay in the arrival of generic entry that results from such a settlement is the source of the other common term "pay for delay." Notice also that this tactic allows the brand to leverage what might be quite weak patents into monopoly profit.

IV. PROBABILISTIC PATENTS AND THE SUPREME COURT

What is a weak patent? It is one that has a low probability of being found valid by a court. The notion of “probabilistic patents” was introduced into the law and economics literature by Carl Shapiro in various articles over the last decade, and has significantly changed economic thinking concerning the way intellectual property works.² The 2013 Supreme Court pay-for-delay decision (*FTC v Actavis*) makes it clear that the important insights in those pieces have only begun to be integrated into legal reasoning.

A patent, when issued by the PTO, may or may not be valid. It has some probability $p \leq 1$ of being found valid in litigation. The statistics on patent validity in the US unfortunately suggest that p is, on average, quite small. The PTO issues 15,000 patents each month and each application receives 15-to-20 hours of patent examiner time on average.³ Of patents that are litigated to trial (0.1 percent), approximately half are found invalid.⁴ Between 55 and 67 percent of patents are not renewed with user fees, suggesting that their owners do not believe they are worth retaining, perhaps because of invalidity or perhaps because of commercial irrelevance.⁵

The Supreme Court pay-for-delay decision revisits a case where the Federal Trade Commission sued Actavis under the theory of harm described above. A district court dismissed the case, and the Eleventh Circuit affirmed, relying on the public policy favoring settlement of disputes and stating that the patent owner was contracting within the scope of the patent.

THE PARTIES PREFER TO SETTLE PATENT LITIGATION IN THIS MANNER BECAUSE IT PRESERVES THE BRAND'S MONOPOLY PROFIT FOR DIVISION BETWEEN THE TWO FIRMS; THE MISSING ELEMENT, OF COURSE, IS THAT CONSUMER SURPLUS IS NOT CONSIDERED IN THE ARRANGEMENT. THE PROFIT ACCRUING TO BOTH PARTIES WOULD BE SHARED WITH CONSUMERS TO A SIGNIFICANT DEGREE IF THE USUAL ENTRY PROCESS TOOK PLACE.

First consider the competitive situation in the event the brand has a known-to-be valid patent. Such a patent would not attract a generic challenge. In the event of a challenge, the most the brand would be willing to pay to settle would be exactly the cost of its litigation because both the brand and the generic know that its patent is valid. Leaving aside litigation costs, in such a setting a reverse payment will never be used. Thus we see that the Eleventh Circuit's reasoning contains an internal contradiction: if the patent is valid, no reverse payment should exist. Indeed, the presence of a reverse payment itself suggests the patent may not be valid.

The majority opinion of the Supreme Court contains an important subtlety:

“...and we are willing to take this fact as evidence that the agreement's anticompetitive effects fall within the scope of the exclusionary *potential* of the patent.”⁶

The key difference in the conclusions of the two courts, in my view, comes from the interpretation of the word “potential.” We know from the empirical evidence that, if litigated, many patents do not have much exclusionary power. Thus the “exclusionary potential” of a patent might be quite small. Implicitly, the reasoning in the majority opinion is that the agreement's anticompetitive effects (the elimination of generic competition) may not fall within the proper scope of a weak patent. The majority observes that the reverse payment allows

the patent holder effectively to buy the remaining validity it has not earned through its own innovation:

“The payment in effect amounts to a purchase by the patentee of the exclusive right to sell its product, a right it already claims, but would lose if the patent litigation were to continue and the patent were held invalid or not infringed by the generic product...”⁷

The majority is also quite clear on the large competitive harm from this behavior.

“But settlement on the terms said by the FTC to be at issue here—payment in return for staying out of the market—simply keeps prices at patentee-set levels, potentially producing the full patent-related \$500 million monopoly return while dividing that return between the challenged patentee and the patent challenger. The patentee and the challenger gain; the consumer loses.”⁸

The dissent by Chief Justice John Roberts relies heavily on the scope of the patent reasoning; it is the main argument he uses to find in favor of Solvay. The dissent’s argument assumes that the patent is valid and infringed with $p=1$. Of course, in that environment, excluding a competitor is appropriate. However, we know that patents with $p=1$ are rare, and that the average patent has a significant chance of being found invalid. This fact is never raised in the dissent. Rather, it returns again and again to the issue of the scope of the patent without appreciating its probabilistic nature. However, the arguments the dissent makes are exactly right once the probabilistic nature of patents is taken into account. This small change substantially reconciles the two opinions.

“The correct approach should therefore be to ask whether the settlement gives Solvay monopoly power beyond what the patent already gave it.”⁹

Interestingly, though it is stated as a criticism, this statement is exactly what the majority is doing. The difference is that in the US in the modern era, we cannot assume an unlitigated patent is valid. That is not a position that is empirically correct. The only way to view patents that is consistent with the data is as probabilistic rights.¹⁰ The majority is acting consistently with the premise that the *expected* validity of any given patent is less than one, and that the patent at issue in this case might have a very low probability of being valid. If the patent has a low probability of being valid, the settlement may give the patent owner monopoly power well beyond what the patent provides.

The dissent goes on to discuss the notion of the zone within which the patent holder may operate as being critical to the application of antitrust law.

“The point of patent law is to grant limited monopolies as a way of encouraging innovation... In doing so it provides an exception to antitrust law, and the scope of the patent—*i.e.*, the rights conferred by the patent – forms the zone within which the patent holder may operate without facing antitrust liability.”¹¹

“The key, of course, is that the patent holder—when doing anything, including settling—must act within the scope of the patent. If its actions go beyond the monopoly powers conferred by the patent, we have held that such actions are subject to antitrust scrutiny.”¹²

“...that when a patent holder acts outside the scope of its patent, it is no longer protected from

antitrust scrutiny by the patent.”¹³

These quotations from the dissent are exactly right—and consistent with the majority opinion. For example, if the patent is valid with $p=.5$, its scope is limited; the rights holder could settle for an entry date that is half the patent life to achieve all the profit to which it is entitled. Settling with the generic to achieve the full patent life would be stepping outside the “zone” of the rights conferred by the patent and, by these arguments, exposes the patentee to antitrust liability. In a world of probabilistic patents, antitrust enforcement requires that the patent holder not leverage his $p<1$ into $p=1$.

However, it is clear that the dissent rejects the idea that patents could be probabilistic and their scope thereby limited. Below, I reproduce part of the majority opinion that Chief Justice Roberts quotes in disagreement, and then the dissent’s own conclusion:

“First, the majority explains that ‘the patent here may or may not be valid, and may or may not be infringed’ Because there is ‘uncertainty’ about whether the patent is actually valid, the Court says that any questions regarding the legality of the settlement should be ‘measure[ed]’ by ‘procompetitive antitrust policies’ rather than ‘patent law policy.’”¹⁴

“And the scope of the patent – *i.e.* what rights are conferred by the patent – should be determined by reference to patent law.”¹⁵

From this it appears that Chief Justice Roberts thinks that determining legality using “patent law policy” means that the judge or enforcement agency should assume all patents are valid, in contrast to the empirical evidence. Notice that such a policy endows a party that files what could be a completely useless patent with tremendous market power. Indeed, such a policy encourages the filing of trivial patents because there is a chance they can be used to prolong and maintain monopoly profits.

Finally, I will address the issue of why such reverse payments have been rare in the past. (“The majority points to no case where a patent settlement was subject to antitrust scrutiny merely because the validity of the patent was uncertain. Not one.”)¹⁶ Hatch-Waxman sets up unique incentives that cause reverse payments to be worth making. In a standard market where entry is not regulated by the government, there is no way to stop attempted entry by every possible entrant. Normally, as the majority opinion points out, a rights holder would have to pay off a never-ending line of patent challengers. In this setting a patent holder would not find reverse payments to be efficacious in maintaining its monopoly. By contrast, HW endows that first generic Paragraph IV challenger with the ability to block others from entering. The regulation itself creates only one partner with this power and therefore one party with whom it is

THE ONLY WAY TO VIEW PATENTS THAT IS CONSISTENT WITH THE DATA IS AS PROBABILISTIC RIGHTS.¹⁰ THE MAJORITY IS ACTING CONSISTENTLY WITH THE PREMISE THAT THE EXPECTED VALIDITY OF ANY GIVEN PATENT IS LESS THAN ONE, AND THAT THE PATENT AT ISSUE IN THIS CASE MIGHT HAVE A VERY LOW PROBABILITY OF BEING VALID. IF THE PATENT HAS A LOW PROBABILITY OF BEING VALID, THE SETTLEMENT MAY GIVE THE PATENT OWNER MONOPOLY POWER WELL BEYOND WHAT THE PATENT PROVIDES.

HATCH-WAXMAN SETS UP UNIQUE INCENTIVES THAT CAUSE REVERSE PAYMENTS TO BE WORTH MAKING.

IN A STANDARD MARKET WHERE ENTRY IS NOT REGULATED BY THE GOVERNMENT, THERE IS NO WAY TO STOP ATTEMPTED ENTRY BY EVERY POSSIBLE ENTRANT.... BY CONTRAST, HW ENDOWS THAT FIRST GENERIC PARAGRAPH IV CHALLENGER WITH THE ABILITY TO BLOCK OTHERS FROM ENTERING.

worth colluding. It is not surprising that we have not seen many patent owners try this tactic before HW, nor that quite a few have tried it since.

VI. POLICY GOALS

The Supreme Court decision finding that reverse payments should be subject to the rule of reason allows antitrust enforcers to pursue many admirable policy goals. One important policy goal is to create an environment where innovation is rewarded with patent rights. Secondly, it seems desirable to encourage pro-competitive and legitimate market interactions among firms, such as JVs, supply agreements, co-marketing,

etc., even if the parties are a firm that makes brand and a firm that makes generics. However, policy makers have strong reasons to promote generic entry into a market as soon as intellectual property rights permit. Competition enforcement can help prevent the brand from blocking entry with legal strategies that bolster its market power beyond the value of its innovation.

The FTC proposal to encourage settlement negotiations over entry date, but not money, has good incentive features. The idea is that instead of allowing the brand to compensate the generic to stay off the market, the parties may only bargain over the date the generic may enter. In that setting, the generic wants to enter early (because that is the only way it can earn profits) while the brand wants the generic to enter late (to preserve its own profits). The strength of the patent will be key in determining the final negotiated date that settles the litigation. In this way the monopoly power that results from the settlement is exactly consistent with the patent's scope, and consumers benefit from competition when that scope ends. ▲

REFERENCES:

Farrell, Joseph and Carl Shapiro (2008) "How Strong are Weak Patents" *American Economic Review*: 98:4:1347-1369.

Hemphill, Scott (2006) "Paying for Delay: Pharmaceutical Patent Settlements as a Regulatory Design Problem" *NYU Law Review*, Vol. 81, p1553.

Lemley, Mark and Carl Shapiro (2005) "Probabilistic Patents" *Journal of Economic Perspectives*: 19:2:75-98.

Shapiro, Carl (2003) "Antitrust Limits to Patent Settlements" *RAND Journal of Economics*: 34:2:391-411

1. Professor of Economics, Yale University
2. See, for example, Farrell and Shapiro (2008) Lemley and Shapiro (2005), and Shapiro (2003)
3. Farrell and Shapiro (2008) p 1347
4. Lemley and Shapiro (2005) p75-6
5. Lemley and Shapiro (2005) p80
6. Fed. Trade Comm'n v. Actavis, Inc., 133 S. Ct. 2223, 2230 (2013) (emphasis added).
7. Id. at 2234.
8. Id. at 2234-35.
9. Id. at 2238 (Roberts, C.J., dissenting).
10. Lemley and Shapiro (2005)
11. Id.
12. Id. at 2239.
13. Id. at 2241.
14. Id. at 2239-40.
15. Id. at 2240.
16. Id. at 2242.



*Current Cases
and Issues*

There is Always a First Time:

Coordinated Effects via Vertical Structural Changes in Anglo/Lafarge

BY JULIE BON, PIETRO CROCIONI & FRANCESCA SALA¹

This article explores the UK Competition Commission's Anglo/Lafarge merger decision (2012) focusing on the reasoning for a finding of coordinated effects in cement. While the theory of harm identified has been considered in a few theoretical academic papers and mentioned in the European Commission's Guidelines on Non-Horizontal Mergers (2008), to our knowledge this was the first SLC finding in Europe based on coordinated effect in which the vertical aspect of the merger was a critical component of the decision.

I. INTRODUCTION

Mergers that raise coordinated effects concerns are a rare occurrence in Europe. Those that raise such concerns through changes in the vertical structure rather than more simply increasing concentration or eliminating a maverick firm are even rarer. *Anglo/Lafarge* was such a merger (though technically a joint venture). It was even more unusual given that the merger was between two cement producers that were already vertically integrated in downstream ready-mix (RMX) production—cement is a key input to produce RMX—pre-merger and, hence, one that did not affect the market share of the independent (*i.e.* not vertically integrated) downstream RMX providers. Instead, the merger increased the degree of vertical integration of one provider and critically resulted in cement producers being more similar in terms of their degree of vertical integration.

The Anglo/Lafarge merger (CC, 2012) was not solely a “vertical” merger. It also substantially increased

MERGERS THAT RAISE COORDINATED EFFECTS CONCERNS ARE A RARE OCCURRENCE IN EUROPE. THOSE THAT RAISE SUCH CONCERNS THROUGH CHANGES IN THE VERTICAL STRUCTURE RATHER THAN MORE SIMPLY INCREASING CONCENTRATION OR ELIMINATING A MAVERICK FIRM ARE EVEN RARER. *ANGLO/LAFARGE* WAS SUCH A MERGER

concentration upstream in cement production and led to consolidation in the downstream RMX markets in the UK. However, the Competition Commission's (CC) finding that the merger would have led to a Significant Lessening of Competition (SLC) in the UK cement market because of coordinated effects in the production of cement was not based solely on the impact of increased concentration in cement on coordination. The CC also found that the increased vertical integration would have had a substantial impact on the likelihood of coordination post-merger. Indeed, the CC found that simply divesting cement capacity

equal to the incremental increase brought about by the merger was, on its own, insufficient to allay concerns about coordinated effects.

This was not only a coordinated effects merger. The overall merger assessment was made more complex by also raising unilateral effects concerns in the supply of rail ballast and high purity limestone and, locally, in primary aggregates, RMX and asphalt.

This article explores the CC's *Anglo/Lafarge* decision (2012), focusing solely on the reasoning for a finding

of coordinated effects in cement. To our knowledge this is the first merger decision in Europe with an SLC finding of coordinated effects based in part on a change in the vertical market structure.

This article is organized as follows:

- Section 2 discusses the circumstances in which vertical mergers can facilitate coordination. It briefly reviews the relevant literature, the main competition authorities' guidelines and some potentially relevant merger cases;
- Section 3 describes and examines the Anglo/Lafarge decision and briefly considers how it relates to the existing literature; and
- Section 4 offers some concluding remarks.

II. WHEN CAN VERTICAL MERGERS FACILITATE COORDINATION?

We examined the relevant literature, the EU and US competition authorities' merger guidelines and selected cases to derive a simple taxonomy of how economists currently consider vertical mergers could facilitate coordinated behavior upstream.

It is not straightforward to define a vertical merger. The simplest example is that of an upstream firm merging with a downstream firm. This increases vertical integration in the market by reducing the number of non-integrated downstream firms. However, the Anglo/Lafarge merger did not create a new vertically integrated producer as the two merging firms were already vertically integrated pre-merger, both being active in the upstream input market—cement—and in the downstream product market—RMX—although to substantially different degrees. Tarmac (owned by Anglo-American) almost entirely self-supplied its RMX plants with cement, while Lafarge, with a smaller presence in RMX, sold a substantial proportion of its cement to independent RMX providers. However, the CC found that, over and above the increased upstream concentration, the fact that the merger increased Lafarge's downstream presence was likely to further facilitate coordination. In other words, this merger was vertical in the sense that it increased the degree of vertical integration of an already vertically integrated provider (Lafarge). In addition, by so doing, it increased the similarities of the post-merger cement producers in terms of their degree of vertical integration. In some circumstances, this could have important implications as to the ability and sustainability of coordination.

A. Literature

The literature that examines how vertical mergers may facilitate collusion is recent and very limited in scope. The analysis of vertical of mergers has, instead, almost entirely focused on foreclosure rather than coordinated effects.

There are two notable exceptions: the contributions by Nocke and White (2007) and Norman (2009).

Nocke and White (2007) examines the impact of a vertical merger on the sustainability of collusion—*i.e.* what is the impact of a merger between an upstream and a downstream firm on the sustainability of upstream collusion. The paper identifies two effects from an increase in vertical integration: an “outlet effect,” which facilitates collusion, and a “punishment effect” that undermines it. The outlet effect reduces the profitability of

deviations and, hence, it enhances collusion. This arises because a vertical merger reduces the number and share of supply of non-integrated downstream firms. Upstream rivals who deviate cannot expect to win sales from the downstream division of the vertical integrated supplier anymore. The deviator has then fewer outlets it can hope to win over. The profits from deviations and the incentive to deviate are, hence, reduced. The punishment effect works in the opposite direction and increases the incentives of the vertically integrated firm to deviate. This is because the vertically integrated firm cannot be punished as severely as if it were not vertically integrated as its downstream division is less dependent on external upstream suppliers and will continue to purchase from the upstream division during a punishment phase.

Norman (2009) builds on Nocke and White (2007) by assuming that the upstream input is sold at linear prices rather than using two-part tariffs. It confirms Nocke and White's results. With linear prices, the upstream firms cannot realize the maximum industry profit because there will be double marginalization. This has important consequences as to the overall welfare implications because, although a vertical merger may facilitate collusion upstream, this may not lead to a decrease in welfare. In fact, the vertical merger may have the effect of eliminating the double mark-up, which may offset the welfare effect of increased upstream coordination.

The Jullien and Rey (2007) article does not deal directly with the impact of vertical mergers. However, the authors argue that Retail Price Maintenance (RPM) can help collusion by making it easier for upstream firms to monitor each other's behavior. If the downstream market were subject to demand or supply shocks, downstream firms would reflect such shocks by adjusting their prices. RPM removes this price flexibility and therefore lowers any collusive profits. However, it also increases the ability to detect deviations by *de facto* eliminating the possibility that downstream prices changed due to anything by deviations. The overall effect of RPM on collusion is ambiguous; however, if RPM has no efficiency features, then the collusion facilitating effect is likely to dominate. Vertical integration can provide a similar effect as RPM in terms of increased price visibility and, hence, better ability to detect deviations. Therefore, by integrating downstream, firms would be better able to monitor the behavior of their upstream rivals.

B. Merger Guidelines

The European Commission and the US Department of Justice and Federal Trade Commission merger guidelines discuss the possible impact of vertical mergers on coordination.²

The European Commission Non-Horizontal Merger Guidelines³ put forward the three necessary conditions for coordination to be sustainable:

- the coordinating firms must be able to monitor to a sufficient degree whether the terms of coordination are being adhered to;
- discipline requires that there is some form of deterrent mechanism that can be activated if deviation is detected; and
- the reactions of outsiders, such as current and future competitors not participating in the coordination, as well as customers, should not be able to jeopardize the results expected from the coordination.⁴

They also stress how a vertical merger could impact on all the three necessary conditions for coordination.

It could increase the firms' ability to reach a common understanding and to monitor competitors' behavior by, for example, leading to foreclosure and reducing the number of competitors, increasing the symmetry of firms, or eliminating a maverick firm.⁵ It could also increase the ability to monitor deviations by increasing transparency, if this is higher downstream than upstream, and firms' ability to punish deviators.⁶ It may also increase the external stability of any coordination by raising barriers to entry or eliminating a disruptive buyer.⁷

The US Non-Horizontal Merger Guidelines⁸ refer to two circumstances in which vertical integration could facilitate collusion. First, increased vertical integration may facilitate upstream collusion by making it easier to monitor price changes. This is because retail prices are generally more visible than prices in upstream markets, and vertical mergers may increase the level of vertical integration to the point at which the monitoring effect becomes significant. The US Guidelines conclude that concerns would be unlikely unless the upstream market is generally conducive to collusion and most of its sales are via vertically integrated retail outlets. Second, a vertical merger may eliminate a particularly disruptive buyer in a downstream market and may thus facilitate collusion upstream. Such a buyer may disrupt collusion if it is sufficiently important to induce a seller to deviate from the terms of a collusive agreement in an effort to secure that business. The merger of such a buyer with an upstream firm may eliminate that incentive and make collusion more likely or effective.

C. *Relevant Merger Cases*

We have also briefly reviewed selected vertical merger cases that have raised coordinated effects concerns at the European level.⁹ Although the European Commission has raised serious concerns in a few concentrations because of coordinated effects concerns,¹⁰ it has not yet done so in vertical mergers for the same reason.

Two almost contemporaneous vertical mergers—*TomTom/Tele Atlas* (2007)¹¹ and *Nokia/Navteq* (2008)¹²—involved the supply of digital navigation services. Digital maps are the essential upstream input in order to provide customers with navigation services. Tele Atlas and Navteq were prominent providers of navigable digital maps. They supplied manufacturers of PNDs (Portable Navigation Devices), car manufacturers, navigation software producers, mobile handset manufacturers and location web companies with the digital maps they needed to provide navigation services. Both transactions raised potential input foreclosure concerns, although the two transactions had only a limited impact on each other in terms of competitive assessment, as TomTom and Nokia are essentially active in different downstream markets. The European Commission dismissed foreclosure concerns but also briefly examined any impact on coordination upstream.

The European Commission examined whether the TomTom/Tele Atlas merger was likely to make upstream coordination between Tele Atlas and Navteq likely.¹³ It concluded that this was not the case, as reaching an agreement would be difficult because the prices of map databases were not transparent and large and infrequent contracts would undermine coordination. The European Commission, though, did not examine whether vertical integration increased Tele Atlas and Navteq's ability to monitor each other's behavior. It also concluded that the Nokia/Navteq merger was not likely to lead to coordinated effects in the upstream market for navigable maps.¹⁴ Despite noting that there was evidence of vibrant competition pre-merger, the European Commission also noted that, following the two mergers, there would be two vertically integrated providers (Nokia/Navteq and TomTom/Tele Atlas). Although TomTom and Navteq would have a common incentive to partition the market between the supply to PND and mobile devices, the European Commission found that such a strategy would be unlikely because the growth expectations for the two devices were very different. This

would make any market partition along these lines inherently unstable. Building on the analysis in TomTom/Tele Atlas, the European Commission also concluded that the vertical integration of Nokia and Navteq would not increase price transparency.

In the *Accor/Hilton/Sixcontinents* joint venture, the European Commission (2003)¹⁵ also examined whether this vertical merger between hotel operators on the one hand (Accor, Hilton and Six Continents) and Worldres on the other (a provider of business-to-business e-commerce solutions for online marketing and reservations to hotel operators) would raise coordinated effects concerns. In particular, the European Commission considered whether hotel operators would be able to coordinate via sharing of the information stored and processed by Worldres. The European Commission found that the merger was not likely to lead to coordination via such a mechanism.

THE ANGLO/LAFARGE MERGER IS AN EXAMPLE OF A CHANGE IN THE DEGREE OF VERTICAL INTEGRATION IN THE CEMENT AND RMX MARKETS THAT WAS FOUND TO FACILITATE UPSTREAM COORDINATION IN THE SUPPLY OF CEMENT, ALTHOUGH THERE ARE ALSO IMPORTANT EFFECTS OF UPSTREAM INCREASED CONCENTRATION ON COORDINATION IN CEMENT.

D. A Taxonomy

From our brief review of the relevant literature and the merger guidelines we identified a number of ways in which vertical mergers could facilitate collusion. A useful categorization would be to distinguish cases depending on whether the coordination concern would arise upstream or downstream. When the impact of the merger is to solely increase concentration upstream (or downstream) and coordinated effects concerns relate to the same market, the merger has no vertical impact on coordination.

Of the two cases where there is a vertical effect, the one most often considered is that of a merger that facilitates coordination upstream. There are a number of possible explanations as to how a vertical merger could facilitate upstream coordination:

- The “outlet effect” discussed by Nocke and White (2007);
- The increased ability to monitor and punish upstream competitors’ behavior mentioned in the EU and US Non-horizontal Merger Guidelines; and
- Elimination of a disruptive/maverick downstream buyer discussed in the US Non-horizontal Merger Guidelines.

The Anglo/Lafarge merger is an example of a change in the degree of vertical integration in the cement and RMX markets that was found to facilitate upstream coordination in the supply of cement, although there are also important effects of upstream increased concentration on coordination in cement. The discussion in the next section explains the reasoning in the CC’s decision on coordinated effects in the Anglo/Lafarge merger inquiry.

III. THE ANGLO/LAFARGE MERGER

This joint venture would have combined the UK activities of Tarmac (a subsidiary of Anglo American) and Lafarge. They planned to contribute to the joint venture their UK activities in the production of cement, aggregates, asphalt and RMX.

Given the focus of this paper on coordinated effects in (bulk) cement, the relevant activities of the merger parties, in this context, are in cement and RMX. Of the 12 cement plants in Great Britain¹⁶ at the time of the CC's inquiry, Tarmac operated one plant (Tunstead). It also operated over 150 fixed RMX plants. Lafarge operated four cement plants (Hope, Cauldon, Dunbar and Aberthaw), 12 cement depots and two cement import terminals.¹⁷ It also had about 90 fixed RMX plants. In broad terms, Tarmac self-supplied its RMX plants with the cement it produced, while Lafarge's RMX's operations were more limited relative to its cement production and Lafarge sold a large proportion of its cement to third parties. There are two other cement producers in Great Britain, Hanson and Cemex, both of which are vertically integrated into RMX. Hanson and Cemex have a similar degree of vertical integration to each other, and it lays in between that of Lafarge on one hand (which had a relatively low degree of vertical integration) and Tarmac on the other (which had a relatively high degree of vertical integration). Most cement producers also buy and sell cement to each other, although this has become less common more recently.

The relationship between cement and RMX is critical in understanding the vertical aspect of the CC's coordinated effects finding. Cement is used to bind together the components of building materials. It is therefore mixed with aggregates and water to produce RMX or concrete products such as prefabricated building blocks made out of concrete. Cement and aggregates are therefore the two key inputs into the production of RMX.

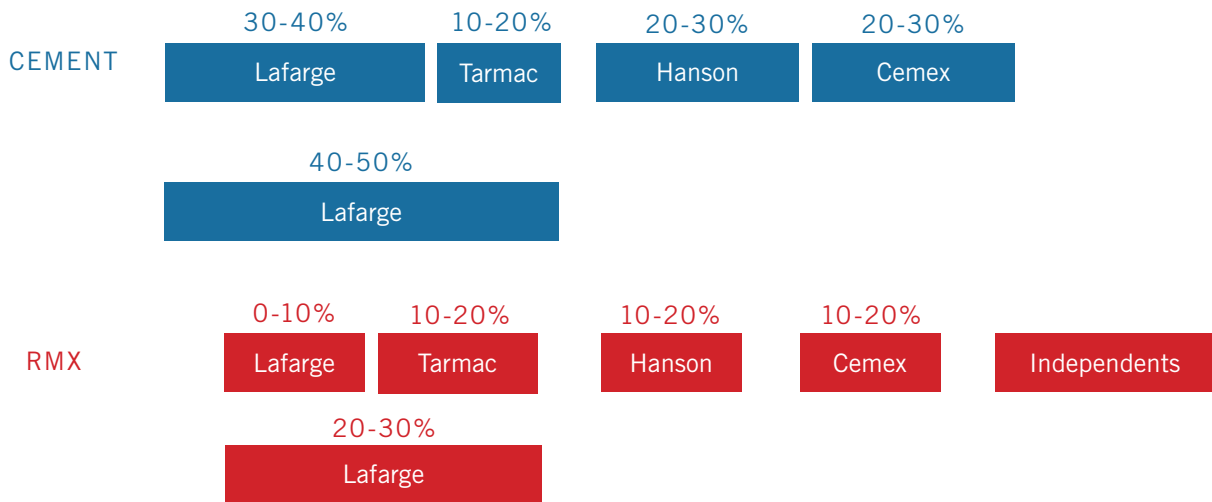
A. *Structural Impact of the Merger*

The CC concluded that the merger would have three broad structural effects that were relevant for the assessment of coordinated effects:¹⁸

- increased concentration in cement production;
- increased concentration in RMX production; and
- a more balanced position in terms of the degree of vertical integration (cement vs. RMX) between the cement producers.

This is illustrated in Figure 1, which shows that as a result of the merger, the new joint venture (JV) would, in comparison with Lafarge, self-supply a much larger proportion of its cement and rely less on sales to third parties. The JV would therefore become much similar to the two remaining vertically integrated cement producers, Hanson and Cemex.

Figure 1: Structure of the (bulk) cement and RMX markets in Great Britain: market shares of bulk cement sales and shares of RMX sales (2010)



Source for market share ranges: CC Anglo/Lafarge Decision (2012), Table 10. The boxes are not in scale.

B. Pre-Existing Coordination

The CC examined market characteristics and outcomes and assessed whether the three conditions for coordination identified in the CC/OFT Merger Assessment Guidelines (2010) were met in the Great Britain cement market pre-merger¹⁹ and, in particular:

- the cement producers' ability to reach and monitor coordination;
- the internal sustainability of such coordination; and
- its external sustainability.

It found that the evidence was consistent with a degree of pre-merger coordination on the basis of a range of evidence including:

- pricing behavior and sustained margins that did not appear to be consistent with the excess capacity in the industry. In particular, increases in the variable profits per ton of cement over the period 2007 to 2010 appeared inconsistent with cement producers competing for customers in a market where demand fell substantially and led to excess capacity;
- the degree of stability of shares of production at the time of large changes in demand and in the structure of the industry; and
- the results from the CC's econometric estimation which may be consistent with the existence of a degree of coordination in the market.²⁰

The CC concluded that if there was pre-existing coordination, it would have been between all Great Britain cement producers with the exception of Tarmac. Coordination would have operated by monitoring shares of

total cement production and/or wins and losses of customers, rather than prices (which are individually negotiated with customers and lack transparency). Producers would also signal their future intentions by:

- issuing generic price announcement letters to customers and monitoring of others' price announcements; and
- reverting to self-supplying small volumes of cement that were previously purchased from another producer (this action is known as "repatriation"). This would act as an additional cheap signal to potential deviators to stop current deviations, without necessarily getting into costly retaliatory actions.

Deviations could have been punished by lowering cement prices to independent RMX producers, or by reducing RMX prices charged by integrated RMX businesses to RMX customers.²¹

The CC also found that any pre-existing coordination in the bulk cement market would not be undermined by external factors (*i.e.* importers, entry and countervailing buyer power).

C. The Impact of the Merger

The CC did not conclude on whether there was pre-existing coordination in the supply of cement. This is not a requirement in order to find that a merger would lead to coordinated effects. Rather, the CC analyzed the impact of the merger under two different scenarios for the pre-merger situation: the case where there was some pre-existing coordination and the case where there was no pre-existing coordination. In this article we principally focus on the case in which there is no pre-merger coordination, but we first briefly examine how the merger may have strengthened any pre-existing coordination.

...POST-MERGER CEMENT PRODUCERS WOULD BE BETTER ABLE TO MONITOR COORDINATION, AS THEY WOULD BE ABLE TO SPOT DEVIATIONS OR TO TARGET PUNISHMENT MORE EASILY WITH ONE FEWER PRODUCER. MONITORING WOULD BE LIKELY TO BE EASIER AND ANY INFERENCES DRAWN FROM OBSERVING A REDUCTION IN CEMENT SALES WOULD BE MORE PRECISE (*E.G.* DISTINGUISHING BETWEEN A DEVIATION AND A CHANGE IN DEMAND FOR CEMENT).

We focus on the effect of the merger on the three conditions for coordination and the impact on all cement producers but with a particular emphasis on how the merger would have altered the structure of Lafarge's cement and RMX business. Although post-merger Lafarge's and Tarmac's would have been conferred into a JV, for ease of exposition we refer to Lafarge pre and post-merger. We analyzed whether the merger either strengthened the conditions for coordination (if there was already evidence of coordination pre-merger) or whether the impact of the merger created the conditions for coordination (absent pre-merger coordination).

D. If There Were Pre-Existing Coordination

The CC concluded that the merger would make any pre-existing coordination (working as discussed above) more stable.²²

If there were already a degree of pre-existing coordination, post-merger cement producers would be better able to monitor coordination, as they would be able to spot deviations or to target punishment more easily

with one fewer producer. Monitoring would be likely to be easier and any inferences drawn from observing a reduction in cement sales would be more precise (*e.g.* distinguishing between a deviation and a change in demand for cement). This is because, pre-merger, Tarmac was not involved in coordination and its incentive and ability to continue to expand output in the medium term still introduced some uncertainty. This would be removed by the merger.

If there were pre-existing coordination, then post-merger, this would become more stable and, hence, more internally sustainable. Lafarge's post-merger absolute profits from coordination in cement would be larger than those pre-merger, reducing its incentives to deviate post-merger. As a larger producer, any deviation by Lafarge would have a greater impact on the market and be more likely to provoke punishment by the other cement producers.²³

In assessing the effect of the proposed merger on the external sustainability of coordination, the CC focused mainly on the role of Tarmac in the bulk cement market. Pre-merger it found that Tarmac had limited incentives to coordinate compared to the other Great Britain cement producers and therefore it was likely to be part of a competitive fringe.²⁴ Furthermore, Tarmac could not expand its cement sales further in the short term, given that it was at, or close to, full capacity. However, Tarmac had considerably expanded its capacity on two occasions in the ten years prior to the inquiry. It may, therefore, have been perceived as a long-term potential threat to any possible coordination.²⁵ Post-merger, the threat that Lafarge might expand its capacity further would be lower as Lafarge's cement plants had excess capacity pre-merger.

E. Absent Pre-Existing Coordination

The CC then examined the impact of the merger on each of the conditions for coordination in the case where there was no pre-existing coordination.

1. ABILITY TO REACH AND MONITOR COORDINATION

The merger would raise concerns if it made it more likely that Great Britain cement producers could reach a common understanding on the terms of coordination, or that they could do so more easily.²⁶

The CC considered two factors that might facilitate reaching and monitoring coordination post-merger.

First, the number of cement producers would decline from four to three making it easier and/or quicker to reach a common understanding on the terms of coordination. Furthermore, post-merger, cement producers would also have an increased incentive to reach an understanding of the benefits they could achieve from coordination. This is because, for example, any act of aggressive competition by a cement producer would affect the remaining two competitors more strongly than pre-merger.

Second, Lafarge would have access to more information about the market post-merger. Its increased degree of vertical integration and, in particular, its larger presence in RMX, would provide it with more information to facilitate upstream coordination. Post-merger Lafarge would have a better understanding of the RMX market (both overall and in its geographic distribution), via informal local contacts with RMX purchasers rather than having to observe competitors' cement sales volumes in these areas. In other words, additional RMX plants

would provide better local knowledge to spot deviation in cement. For example, pre-merger, in areas where Lafarge did not have RMX plants, it might find it difficult to distinguish between when its cement sales declined because of an overall decline in demand in the area and when they declined because of deviations. The incremental RMX plants gained through the merger would change from being, in some cases, a customer that could potentially switch supplier and that, hence, needed to be monitored, to being a vertically integrated retail outlet for the Lafarge's cement post-merger.

The merger parties argued that additional RMX plants would be likely to provide at best only limited incremental transparency since changes in RMX sales volumes were driven by a large number of factors other than the price of bulk cement. However, the basis of the CC's concern was not that Lafarge would have a better ability to monitor RMX sales, but that it would obtain additional information on cement sales, in particular, a better ability to distinguish between the impact on the demand for cement of market changes and of deviations.²⁷

In summary, the CC concluded that, in addition to the effect from the reduction in the number of cement producers as a result of the merger, Lafarge's increased presence in RMX post-merger would be likely to provide it with additional and better information about cement sales. Lafarge would gain general information on the local area via the additional RMX plants.

2. INTERNAL SUSTAINABILITY

In assessing the impact of the merger on the internal sustainability of coordination the CC examined its likely effects on both the incentives to coordinate and the ability to punish deviations.²⁸

The CC considered that the merger would have increased the cement producers' incentives to coordinate because:

- there would be fewer cement producers post-merger. Any coordination would also have been more stable because Lafarge would have stronger incentives to coordinate post-merger because of its larger size, while Hanson's and Cemex's incentives would be unchanged; and
- it would increase the similarities between the cement producers in terms of their degree of vertical integration. Pre-merger Lafarge was the least vertically integrated cement producer with a modest position in RMX. Post-merger it would become more similar to Hanson and Cemex in relative terms. This has important implications for Lafarge's ability to punish deviators, as discussed next.²⁹
- The merger would affect cement producers' range of tools available to punish deviations and their relative effectiveness. Cement producers could punish deviations by engaging in:
 - targeted cement price reductions to external customers—*i.e.* independent RMX and concrete producers. Post-merger the revenues (and profits) from external cement sales would be proportionally less important for Lafarge than pre-merger. Therefore, if it punished a deviation and this led to stronger competition in external sales of cement, Lafarge would be less affected post-merger because of its greater reliance on internal sales.³⁰ This is similar to the outlet effect identified by Nocke and White (2007)³¹ and has the effect of making punishment more credible; and/or

- targeted RMX price reductions by their in-house RMX operations. Pre-merger this punishment mechanism was unlikely to have been as effective for Lafarge as it was for Hanson and Cemex. Pre-merger Lafarge's fewer RMX plants limited its options if it were either to target lower RMX prices at specific locations and/or to punish on a large scale in RMX. The CC noted that, post-merger, if

POST-MERGER, LAFARGE WOULD HAVE A SIMILAR SET OF PUNISHMENT TOOLS WITH A SIMILAR DEGREE OF EFFECTIVENESS AS THOSE AVAILABLE TO HANSON AND CEMEX. AS A RESULT OF THE INCREASED SIMILARITY IN VERTICAL INTEGRATION, EACH REMAINING CEMENT PRODUCER WOULD HAVE A BETTER UNDERSTANDING OF THE ABILITIES AND INCENTIVES OF EACH OF THE OTHER CEMENT PRODUCERS AND WOULD BE BETTER ABLE TO TAKE THESE EXPECTATIONS INTO ACCOUNT IN ITS OWN BEHAVIOR.

Lafarge intended to punish effectively via lower RMX prices, it would have to identify local areas where the deviator either had material RMX sales or sold material quantities of cement to independent RMX producers (which would reduce their cement purchases from the deviator if demand for their RMX declined).

Post-merger, Lafarge would have a similar set of punishment tools with a similar degree of effectiveness as those available to Hanson and Cemex. As a result of the increased similarity in vertical integration, each remaining cement producer would have a better understanding of the abilities and incentives of each of the other cement producers and would be better able to take these expectations into account in its own behavior.

This aspect of the CC's decision was particularly criticized by the merger parties who claimed that co-

ordination that relied on reaching and monitoring the terms of coordination at the RMX level was speculative and would be undermined by the additional complexity of trying to do so. They also argued that punishment of deviations in the cement market via lower RMX prices would be unattractive. They pointed out that even if targeted RMX punishment could be achieved, it would also provide an additional mechanism by which cheating on any coordinated terms could occur – this is similar to the punishment effect identified by Nocke and White (2007), though it applies to a firm rather than to the market. In this respect the CC noted that Lafarge would sell more cement in total post-merger than it had pre-merger, making it more reluctant to deviate because it would have more to lose from deviating than pre-merger when it was a smaller producer.

In summary, post-merger, Lafarge would have greater flexibility and more options in punishment than it had pre-merger and, in this respect, it would be more aligned to the other Great Britain cement producers. This would be likely to make any coordination more stable³² or more likely to emerge post-merger.³³

F. External Sustainability

The CC concluded that post-merger coordination in the bulk cement market would not be undermined by external factors (*i.e.*, importers, entry and countervailing buyer power). Rather, as discussed in the case of pre-existing coordination, the fact that Tarmac would no longer be a possible fringe competitor would increase the external sustainability of any coordinated outcome.

G. Summary

The CC concluded that coordination would become more likely, could be reached more quickly or become more stable post-merger both because of the merger impact on the cement market structure and the increased presence of Lafarge in RMX. This is a critical aspect of the decision and had implications as to the appropriate remedies. Table 1 summarizes the impact on coordination of these two factors.

TABLE 1: THE IMPACT OF THE ANGLO/LAFARGE MERGER

Change in cement market structure	Lafarge's increased presence in RMX
<ul style="list-style-type: none"> • makes reaching a common understanding on the terms of coordination easier and/or swifter, and makes monitoring of wins and losses of cement customers and production volumes easier as a result of the reduction in the number of producers; • reduces Lafarge's incentives to deviate from the coordinated outcome, since it will have larger overall profits from coordination; and • removes Tarmac as a fringe competitor with a strong incentive to produce at capacity and ability to expand its capacity in the future. 	<ul style="list-style-type: none"> • allows more information on the RMX market to flow to Lafarge compared to Lafarge pre-merger, enhancing its ability to monitor coordination; • creates greater similarities in vertical structure among cement producers. This would better align both their incentives to coordinate and their ability to punish deviations, as well as increase Lafarge's flexibility and options in its punishment actions; and • if Tarmac's present cross-sale arrangements remained, this would have given Lafarge increased ability to use repatriation as a cheap signal that it has detected deviation.

The merger would increase transparency in the cement market via Lafarge's increased presence in the RMX market, by means of a mechanism similar to that described in the EU and US Merger Guidelines and in Jullien and Rey (2007). Lafarge's incentives would be modified by an effect similar to the outlet effect identified by Nocke and White (2007). The merger would also increase Lafarge's ability to punish deviations (while leaving other cement producers' ability to punish unaffected) as discussed in the EU Non-Horizontal Merger Guidelines.

The effects arising from the combination of Tarmac and Lafarge's cement businesses and the effects arising from the combination of their RMX businesses were considered both largely independent and cumulative. For this reason the CC required remedies that included (amongst other things) divestment by the merger parties of both a cement plant and a large number of RMX plants.³⁴

IN ESSENCE, THEREFORE, ALTHOUGH THIS IS A NOVEL DECISION, THERE IS NOT MUCH NEW IN THE PRINCIPLE THAT THE MORE THE SUPPLIERS LOOK ALIKE, THE MORE LIKELY THEY ARE TO THINK ALIKE, BEHAVE SIMILARLY AND HAVE SIMILARLY ALIGNED GOALS; THIS, NOTWITHSTANDING THE FACT THAT THE MERGER WOULD HAVE REDUCED THE NUMBER OF CEMENT PRODUCERS IN GREAT BRITAIN FROM FOUR TO THREE.

IV. CONCLUSIONS

To our knowledge the CC's Anglo/Lafarge decision was the first SLC finding in Europe based on coordinated effect in which the vertical aspect of the merger was a critical component of the decision. Although the European Commission's TomTom/Tele Atlas and Nokia/Navteq decisions briefly assessed this as one of the possible theories of harm, which is explicitly mentioned in the EU Non-Horizontal Merger Guidelines, this was not a critical aspect of these decisions. In Anglo/Lafarge this aspect was central in that a divestment of all the cement assets of Tarmac (a single large and efficient cement plant) was deemed insufficient to remedy the SLC finding. The contribution of the Tarmac's RMX plants was a critical aspect of the SLC finding. The greater similarity between the remaining three cement producers in terms of the proportion of cement that they would self-supply to their own RMX plants was deemed central in two main ways. First, additional RMX plants would have provided additional and better information to Lafarge about the behavior of other cement producers compared to what it had before the merger with its limited downstream presence. This would have increased Lafarge's ability to spot deviations. Second, post-merger the additional RMX plants provided Lafarge with the same wider range of punishment tools available to the other Great Britain cement producers pre-merger.

In essence, therefore, although this is a novel decision, there is not much new in the principle that the more the suppliers look alike, the more likely they are to think alike, behave similarly and have similarly aligned goals; this, notwithstanding the fact that the merger would have reduced the number of cement producers in Great Britain from four to three. Although the latter aspect might have been in itself sufficient for a SLC finding, the vertical aspect was critical for the CC to identify a set of effective remedies. ▲

REFERENCES:

CC (2012), Anglo American PLC / Lafarge S.A. merger inquiry report, 1 May, available at <http://www.competition-commission.org.uk/our-work/anglo-american-lafarge>.

CC/OFT (2010), Merger Assessment Guidelines, CC2, available at http://www.ofc.gov.uk/shared_ofc/mergers/642749/OFT1254.pdf.

Jullien, B. and Rey, P., (2007), Resale price maintenance and collusion, *Rand Journal of Economics*, 38(4), 983-1001.

Nocke, V. and White, L., (2007), Do vertical mergers facilitate collusion?, *The American Economic Review*, 97(4), 1321-1339

Normann, T.H., (2009), Vertical integration, raising rivals' costs and upstream collusion, *European Economic Review*, 53(4), 461-480.

1. Julie Bon (julie.bon@cc.gsi.gov.uk) and Francesca Sala (francesca.sala@cc.gsi.gov.uk), UK Competition Commission (CC); Pietro Crocioni (pietro.crocioni@ofcom.org.uk), seconded to the CC from the Office of Communications (Ofcom). Although, we all worked on this merger case while at the CC, the views and opinions expressed in this paper are the sole responsibility of the authors and do not necessarily reflect those of the UK Competition Commission or of its Chairman. The discussion and information in this article is based on publicly available information. We are grateful to all colleagues that worked with us on this inquiry and, in particular, Caroline Wallace for valuable comments on this article. We are also grateful to Massimo Motta for his suggestions. Perhaps needless to say but the responsibility for any remaining error will be efficiently allocated among the three authors.
2. The UK CC/OFT Merger Assessment Guidelines (2010) do not address this possible effect of vertical mergers.
3. European Commission, “Guidelines on the assessment of non-horizontal mergers under the Council Regulation on the control of concentrations between undertakings”, 2008/C 265/07, available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:265:0006:0025:en:PDF>.
4. European Commission Non-Horizontal Merger Guidelines, *para* 81.
5. European Commission Non-Horizontal Merger Guidelines, *paras* 82-85.
6. European Commission Non-Horizontal Merger Guidelines, *paras* 86-88.
7. European Commission Non-Horizontal Merger Guidelines, *paras* 89-90.
8. Department of Justice (DoJ), “Non-horizontal merger guidelines”, 1997, Section 4.22, available at <http://www.justice.gov/atr/public/guidelines/2614.pdf>.
9. This type of mergers is also rare in other jurisdictions. One US example is Masonite (2001), available at http://www.justice.gov/atr/public/press_releases/2001/8754.pdf.
10. None of the decisions, though, led to a prohibition. The merger was either cleared subject to remedies (ABF/GBI, 2008), withdrawn (EMI/Time Warner, 2000), cleared by the European Commission but the decision annulled by the CFI (Sony/BMG, 2004) or blocked by the European Commission but the decision annulled by the CFI (Airtours/First Choice).
11. COMP M.4854 TomTom/Tele Atlas, available at, http://ec.europa.eu/competition/mergers/cases/decisions/m4854_20080514_20682_en.pdf.
12. COMP M.4942 Nokia/Navteq, available at http://ec.europa.eu/competition/mergers/cases/decisions/m4942_20080702_20682_en.pdf.
13. Footnote 12, *paras* 278-283.
14. Footnote 13, *paras* 395-408.
15. COMP M.3101 ACCOR/HILTON/SIXCONTINENTS/JV, available at http://ec.europa.eu/competition/mergers/cases/decisions/m3101_en.pdf.
16. The CC decision focused on the Great Britain Market (i.e., excluding Northern Ireland) because the market structure of Northern Ireland appeared to be different from Great Britain.
17. It also operates a cement plant in Northern Ireland (Cookstown) and one depot.
18. See CC Anglo/Lafarge decision, at *para* 6.205.
19. See CC Anglo/Lafarge decision, at *paras* 6.134-6.202.
20. The CC carried out a price-concentration analysis. Results suggested that the presence of a Hanson or of a Cemex plant within 50 miles had no statistically significant effect on Lafarge’s external sales price for cement. See CC Anglo/Lafarge decision, at *paras* 6.128-6.199 and Appendix H for more details.
21. See CC Anglo/Lafarge decision, at *para* 6.133.
22. A finding of coordination requires that all conditions are satisfied post-merger. If there was pre-existing coordination, it follows that the three conditions for coordination to arise were met pre-merger. In this case, the CC noted that it was not necessary that the merger would increase the extent to which each condition was satisfied in order to reach an SLC finding on the basis of coordinated effects (See CC Anglo/Lafarge decision, at *para* 6.203).

23. Furthermore, we note that in the medium term any additional increase in cement sales in Great Britain (e.g. if demand recovered from the very low point of 2009) would lead to larger profits for the coordinating group of firms which would have weaker incentives to deviate to capture additional sales.
24. The term competitive fringe refers to firms that are not part of the coordinating (or colluding) group and behave as price takers.
25. In the longer term, Tarmac could use its existing planning permission to increase the capacity of its Tunstead cement plant.
26. See CC Anglo/Lafarge decision, at paras 6.213-6.222.
27. The merger parties also argued that pre-merger neither Lafarge nor Tarmac used detailed market information from their in-house RMX operations to monitor cement supplies to RMX producers. The CC noted that, although the possible monitoring mechanisms that it identified focused on external bulk cement sales and Great Britain shares of production, there could also be informal mechanisms for passing on local information from RMX plants. The fact that Lafarge might currently make only limited use of any information from its RMX plants might simply reflect the fact that its RMX network was the smallest among Great Britain cement producers.
28. See CC Anglo/Lafarge decision, at paras 6.223-6.251.
29. The CC also noted that if the merger led to more similar cost among cement producers this would also facilitate the internal sustainability of coordination. The CC examined the variable production costs of each cement producer and found that in recent years Lafarge's cement plants overall have been among the highest-cost plants in the UK, while Tarmac has been the lowest-cost producer based on the audited data to which the CC had access. However, the CC could not reach firm conclusions on this.
30. Lafarge's stronger presence in local RMX markets post-merger would mean that a reduction in cement prices to external RMX producers could also result in a larger loss of sales and profits in its own RMX activities than Lafarge would experience currently. This would tend to reduce its incentive to punish via lower cement prices compared with the pre-merger situation. The CC recognized this effect, but considered its impact limited, first, because a reduction in the price of cement to external RMX producers might not be fully passed through into RMX prices (e.g. if the local RMX market was not very competitive), and, second, because Lafarge could target any reduction in prices to independent RMX producers in specific local areas so as to minimize the effect on its own RMX operations, because RMX markets are local.
31. Nocke and White (2007)'s argument refers to the fact that a vertical merger reduces the downstream share of independent producers. This means that all upstream producers would have reduced incentives to deviate because by doing so their expected incremental sales would be lower than pre-merger. In this case the argument applies to a single firm, Lafarge, because the size of the un-integrated downstream RMX producers was unaffected by the Anglo/Lafarge merger.
32. The scope for increased similarity in vertical integration and its 'stabilizing' impact on the market was noted in several internal documents from the Great Britain cement producers (See CC Anglo/Lafarge decision, at para 6.239).
33. The CC also noted that if post-merger Lafarge were to have greater cross-sales with Hanson and Cemex than Lafarge had pre-merger, then it would have an enhanced ability to use repatriation as a cheap signal to deviators from the coordinated outcome to cease doing so, as an alternative to or a preliminary step before entering a more costly generalised punishment phase. Hence the merger might result in a lower risk for the cement producers of costly price wars than at present. The CC believed that it was possible that Lafarge would have some cross-sales with other GB cement producers.
34. These RMX plants had, together, RMX sales volumes broadly similar to that of Tarmac's RMX business before the merger.



The Classic

An Introduction to Levin, Klevorick, Nelson, and Winter (1987)

BY DAVID S. EVANS

There are several areas of economics where we know much more today than we did 50 years ago as a result of the accumulation of empirical research. More needs to be done, and much remains debated, but economists have made real progress in behavioral economics, labor economics, and financial economics. And in recent years the use of controlled experiments is helping to revolutionize development economics. Even industrial organization has come along. Yet the economics of innovation has not advanced much.

Not from the lack of trying though. There are easily hundreds of published studies of innovation and technical progress. Nonetheless, Wesley Cohen's 2010 summary of *Fifty Years of Empirical Studies of Innovative Activity and Performance* finds few robust findings and ends with a plea for more and better data.¹

According to Cohen, one of the preeminent empirical economists in this area, “we still have little empirical understanding of the tradeoff for industries’ R&D incentives...” and “our understanding of the role of firm-level variables is more primitive still.” Progress has been made on the dynamics of innovation but, “[o]nce again ... the availability of data ... represents an important brake on the advance of empirical testing.” He doesn't quite say this but comes close: one of the problems in this area is that the empirical economists produce results that are hard to explain in the absence of theory, and the theoretical economists produce models that can't be tested with available data.²

There is, unfortunately, a fundamental problem faced by empirical research on the economics of innovation. Labor economists can measure wages, financial economists can measure equity prices, and behavioral economists can observe actual laboratory choices. And they often have pretty good measures of the possible determinants of these outcome variables. Empirical researchers on innovation do not have it so lucky.

No one really knows how to measure innovation. To do studies, economists reach for what they can get. They began with the number of patents. But no one who knows how patents are created could seriously think that patent counts reveal much. Over time economists made this a more meaningful variable by at least adjusting for patent citations under the theory that better patents get more citations. In addition, it is very hard to measure the inputs into innovation. Again economists tend to measure what they can. That could be research and development spending. But companies account for this in different ways and sometimes, particularly for small and young companies, not at all. So the basic problem is that empirical research tries to relate dubious measures of outputs with unreliable measures of inputs.

Cohen calls for better data but he also questions the overwhelming use of econometric models which—my words not his—often in this area are afflicted with the famous garbage-in/garbage-out problem. He observes,

[M]uch of our empirical understanding of innovation derives not from the estimation of econometric models, but from the use of other empirical methods. [T]he historical and case-study literatures provide a rich array of insights and factual information... More strikingly, many

of the most credible empirical regularities have been established not by estimating and testing elaborate models with published data but by the painstaking collection of original data, often in the responses to simple questions.”

That brings me to our classic, *Appropriating the Returns from Industrial Research*, which was authored by four giants of innovation economics: Richard Levin, Alvin Klevorick, Richard Nelson, and Sidney Winter. They tackled one of the core issues in innovation policy: how do firms harvest the fruits of their innovative efforts and how important are patents in enabling this? If they had just run more inconclusive regressions based on publicly available data they would not have created one of the most frequently cited empirical papers on innovation. Instead they asked people in companies who were actually responsible for innovation how they protected the results of their efforts and the relative importance of alternative methods. They conducted their survey across a number of different industries, which then allowed them to say something about the relative importance of different methods for appropriating returns, and the role of patents, in diverse settings.

This research methodology is hardly immune to criticism. People do not necessarily answer questions accurately. Respondents don't use the same metrics, making it difficult to compare responses. Looking at the results, there is tremendous variation even with the same industry. There are many reasons for professional economists to dismiss this effort.

These authors, however, decided not to let the perfect be the enemy of the good. Nor did they resist going against the grain of professional orthodoxy at the time. It paid off. Many of the empirical nuggets—such as the importance of patents in some industries like pharmaceuticals but not in many others—are still quoted. A number of influential papers, some by combinations of the authors of this article and some by others, have built on the original work. The *Appropriating the Returns* authors found that firms did not patent in part because they believed others could invent around them. A subsequent study by Cohen, for example, found that was true but that firms were also concerned about disclosing information in patent filings that could help rivals imitate them.³

The empirical literature on innovation tends towards very narrow analyses of innovation with an almost obsessive focus on the patent process. Levin, Klevorick, Nelson and Winter look at innovation much more broadly. They recognize that there are many ways that firms can protect the fruits of their efforts. These range from trying to get a head start on their rivals, to keeping innovations secret, to providing better service, to patents. If one were going to repeat the study the list would surely be expanded.

The critics and defenders of patents both seem to forget the limited role that patents play in the process of innovation. There are many creations of the mind that are not protected by patents at all. There is some irony in the fact that Albert Einstein could get refrigerator patent for an innovation that has long since been forgotten but a measly Nobel Prize for the photoelectric effect. For all its patents, Apple has not been able to stop the flood of iPad imitators. And somehow innovation happens even when there are no patents available. As Levin and his co-authors showed us, even companies that get patents use many other ways to protect their intellectual property. Patents are but one arrow in the quiver—sometimes a more important, sometimes less important, one. As a result critics and defenders of patents overstate their importance—in both cases because there are imperfect substitutes to which firms would turn if patent rights were eliminated or reduced.

A related point is that too little thought goes into how firms and other institutions such as SSOs would alter

their behavior in the face of changes in patent policy. With even imperfect substitutes available any business that faces a change in one of the inputs into its production process will make other changes in response. The only way to evaluate a change in patent policy is to situate patents, and their various elements, in the broader array of tools for appropriating the returns from R&D.

A bit more than 25 years ago Richard Levin, Alvin Klevorick, Richard Nelson, and Sidney Winter pioneered an important new way of doing empirical research on innovation and produced one of the most durable works in an fragile literature. It is one of CPI's younger but still deserving classics. ▲

1. Wesley M. Cohen, "Fifty Years of Empirical Studies of Innovative Activity and Performance," Chapter 4, Bronwyn Hall and Nathan Rosenberg, *Handbook of Economics of Innovation*, Elsevier 2010, pp. 129-213.
2. In light of the state of the literature on innovation I find it remarkable when economists who venture into policy entrepreneurship rest confident proposals on this very shaky foundation.
3. Cohen, id, 184.

Appropriating the Returns from Industrial Research and Development

BY RICHARD C. LEVIN

Yale University

ALVIN K. KLEVORICK

Yale University

RICHARD R. NELSON

Columbia University

SIDNEY G. WINTER

Yale University

To have the incentive to undertake research and development, a firm must be able to appropriate returns sufficient to make the investment worthwhile. The benefits consumers derive from an innovation, however, are increased if competitors can imitate and improve on the innovation to ensure its availability on favorable terms. Patent law seeks to resolve this tension between incentives for innovation and widespread diffusion of benefits. A patent confers, in theory, perfect appropriability (monopoly of the invention) for a limited time in return for a public disclosure that ensures, again in theory, widespread diffusion of benefits when the patent expires.

ON THE ONE HAND, APPROPRIABILITY IS NOT PERFECT. MANY PATENTS CAN BE CIRCUMVENTED; OTHERS PROVIDE LITTLE PROTECTION BECAUSE OF STRINGENT LEGAL REQUIREMENTS FOR PROOF THAT THEY ARE VALID OR THAT THEY ARE BEING INFRINGED. ON THE OTHER HAND, PUBLIC DISCLOSURE DOES NOT ALWAYS ENSURE ULTIMATE DIFFUSION OF AN INVENTION ON COMPETITIVE TERMS.

Previous investigations of the system suggest that patents do not always work in practice as they do in theory.¹ On the one hand, appropriability is not perfect. Many patents can be circumvented; others provide little protection because of stringent legal requirements for proof that they are valid or that they are being infringed. On the other hand, public disclosure does not always ensure ultimate diffusion of an invention on competitive terms. For example, investments to establish the brand name of a patented product may outlive the patent itself.² And patents may not always be necessary. Studies of the aircraft and semiconductor industries have shown that gaining lead-time and

exploiting learning curve advantages are the primary methods of appropriating returns. Other studies have emphasized the importance of complementary investments in marketing and customer service.³

Evidence on the nature and strength of conditions for appropriability and on the working of the patent system is, however, scattered and unsystematic. Because imperfect appropriability may lead to underinvestment in new technology, and because technological progress is a primary source of economic growth, it would be

useful to have a more comprehensive empirical understanding of appropriability, in particular, to identify those industries and technologies in which patents are effective in preventing competitive imitation of a new process or product. It would also be desirable to know where patents can be profitably licensed. Where patents are not effective, it would be useful to understand why they are not and whether other mechanisms are.

This paper describes the results of an inquiry into appropriability conditions in more than one hundred manufacturing industries. We discuss how this information has been and might be used to cast light on important issues in the economics of innovation and public policy. Our data, derived from a survey of high-level R&D executives, are informed opinions about an industry's technological and economic environment rather than quantitative measures of inputs and outputs.

Although our use of semantic scales to assess, for example, the effectiveness of alternative means of appropriation introduces considerable measurement error; more readily quantifiable proxies would probably not serve as well. Remarkable progress has been made toward developing a methodology to estimate the economic value of patents.⁴ But suitable data are as yet unavailable in the United States, and European data lack sufficiently

reliable detail to support inferences about interindustry differences in the value of patents. Our judgment was that asking knowledgeable respondents about the effectiveness of patents and alternative means of appropriation was at least as likely to produce useful answers as asking for quantitative estimates of the economic value of a typical patent.

INQUIRY INTO APPROPRIABILITY
CONDITIONS IN MORE THAN ONE
HUNDRED MANUFACTURING
INDUSTRIES... DERIVED FROM A SURVEY
OF HIGH-LEVEL R&D EXECUTIVES

BECAUSE TECHNOLOGICAL PROGRESS
IS A PRIMARY SOURCE OF ECONOMIC
GROWTH, IT WOULD BE USEFUL TO HAVE
A MORE COMPREHENSIVE EMPIRICAL
UNDERSTANDING OF APPROPRIABILITY,
IN PARTICULAR, TO IDENTIFY THOSE
INDUSTRIES AND TECHNOLOGIES IN
WHICH PATENTS ARE EFFECTIVE IN
PREVENTING COMPETITIVE IMITATION
OF A NEW PROCESS OR PRODUCT.

We have taken considerable care to establish the robustness of our findings in the presence of possibly substantial measurement error, but ultimately the value of the data will depend on their contribution to better empirical understanding of technological change and more dispassionate discussion of public policy. To view the empirical contribution of the data from the simplest perspective, consider their potential for improving the quality of research that uses patent counts to measure innovative activity.⁵ This line of inquiry has shown, among other results, that industries vary significantly in the average number of patents generated by each dollar of R&D investment.⁶ Our findings on industry differences in patent effectiveness may help explain this variation in the apparent productivity of R&D.

More fundamentally, large and persistent interindustry differences in R&D investment and innovative performance have resisted satisfactory explanation, in part for lack of data that adequately represent the theoretically important concepts of appropriability and technological opportunity. Promising but ultimately unsatisfactory results have been obtained in exploratory work that used crude proxy variables and econometric ingenuity to capture the influence of appropriability and opportunity conditions.⁷ Our desire to provide a stronger basis for this line of inquiry was a prominent motive for our survey research and helped to shape its design.

Finally, gathering better information on the nature and strength of appropriability is particularly timely in view of the prominence of current debates on the adequacy of laws and institutions to protect intellectual property. One impetus for change has been the need to clarify and perhaps strengthen the system of property rights at various new frontiers of technology. Thus, for example, recent legislation has adapted copy right law to protect the rights of the creator of new computer software, a new legal framework has been constructed to protect intellectual property embodied in semiconductor chip designs, and important court decisions and

OUR JUDGMENT WAS THAT ASKING KNOWLEDGEABLE RESPONDENTS ABOUT THE EFFECTIVENESS OF PATENTS AND ALTERNATIVE MEANS OF APPROPRIATION WAS AT LEAST AS LIKELY TO PRODUCE USEFUL ANSWERS AS ASKING FOR QUANTITATIVE ESTIMATES OF THE ECONOMIC VALUE OF A TYPICAL PATENT.

administrative actions have shaped the development of a property rights system in biotechnology.⁸

Another spur to change has been the need to resolve conflicts between the aims of social regulation and the exercise of intellectual property rights. For example, the Drug Price Competition and Patent Term Restoration Act of 1984 extended patent lives of pharmaceuticals to compensate for regulatory requirements that delay the introduction of new drugs.

Intellectual property rights also figure prominently among policy issues milling under the banner of competitiveness. Recent annual reports of the US trade representative have focused on the difficulties US manufacturers encounter in protecting intellectual property rights in foreign markets. The trade bill passed in 1987 by the House of Representatives contains several provisions that increase the scope of protection and the opportunities for relief available to US manufacturers confronted with imports that infringe these rights.⁹ Proposed antitrust legislation, motivated by a concern that courts have kept inventors from reaping rewards that patent laws are intended to provide, stipulates that patent license agreements and similar contracts relating to use of intellectual property “shall not be deemed illegal *per se* under any of the antitrust laws.”¹⁰

To the extent that all this activity attempts to rectify obvious inadequacies in existing institutions, the case for reform appears strong and straightforward. It is easy to deplore the blatant copying of innovative integrated circuit designs, the importation of “knock off” copies of trademarked or patented US products, and the piracy of copyrighted written matter and audio and video cassettes. But reforms may yield unintended consequences. In its simplest form, this concern translates into wariness about Trojan horses: provisions brought into the law by the rhetorical tug of “competitiveness” and “intellectual property” may harbor instruments of protectionism and price fixing. Other potential consequences are subtler but no less important. For example, seemingly uniform adjustments of intellectual property, antitrust, or trade law may affect some industries quite differently than others.

And it should not be taken for granted that more appropriability is better, that better protection necessarily leads to more innovation, which yields better economic performance—higher standards of living, better competitiveness, and so on. Better protection may yield more innovation at the cost of incrementally increasing resources devoted to producing the innovation: the larger prize may merely encourage duplicative private effort to capture it.¹¹ Alternatively, better protection may induce innovation of the wrong kind, or it may buy the innovation by further delaying access to it on competitive terms.¹²

The premise that stronger protection will always improve the incentives to innovate is also open to chal-

lenge. Unimpeded diffusion of existing technology is immediately beneficial not only for consumers but also for those who would improve that technology. Because technological advance is often an interactive, cumulative process, strong protection of individual achievements may slow the general advance. This would not

BECAUSE TECHNOLOGICAL ADVANCE IS OFTEN AN INTERACTIVE, CUMULATIVE PROCESS, STRONG PROTECTION OF INDIVIDUAL ACHIEVEMENTS MAY SLOW THE GENERAL ADVANCE.

occur in a hypothetical world without transaction costs, in which efficient contracts to share information would be made. In reality, however, markets for rights to information are subject to major transactional hazards, and strong protection of a key innovation may preclude competitors from making socially beneficial innovations. The semiconductor industry of the 1950s and 1960s provides an excellent example of rapid progress in a cumulative technology that might have been impossible under a regime that strongly protected intellectual property.¹³

The remainder of this paper discusses our survey instrument, the construction of the sample, and the interpretation of the data, then turn to our findings concerning the effectiveness of patents and other means of appropriating the returns from R&D. The results of related work that employs the survey data to reexamine central questions in the empirical literature on R&D are summarized, and we discuss how our findings might contribute to a more discriminating discussion of patent law, antitrust law, and trade policy.

QUESTIONNAIRE DESIGN AND SURVEY METHODS

The content of our questionnaire was shaped with guidance from the conceptual literature on technological change, empirical literature on the economic impact of the patent system, the work of Mansfield and his associates on imitation costs, and numerous case studies.¹⁴ The questionnaire was aimed at high-level R&D managers with knowledge of both the relevant technology and market conditions. To check the interpretability of the questions and the likely validity and reliability of the responses, we pretested the questionnaire with twelve managers representing diverse businesses.¹⁵

To understand how appropriability differs across industries, we asked each respondent to report typical experiences or central tendencies within a particular industry. Respondents were thus treated as informed observers of a line of business rather than as representatives of a single firm, an approach that encouraged cooperation (they were not placed in the position of possibly divulging practices or policies of their own firms), but led inevitably to heterogeneity in the responses within a given industry.

The questionnaire contained four parts. Parts 1 and 2 concerned appropriability; parts 3 and 4 concerned technological opportunity and technological advance. Questions in part 1 asked about the effectiveness of alternative means of protecting the competitive advantages of R&D, limits on the effectiveness of patents, and ways of acquiring knowledge of a competitors' technology. Part 2 asked about the cost and time required to imitate innovations of rivals; we distinguished process from product innovations, major from typical, and patented from unpatented.¹⁶ Part 3 explored the links between an industry's technology and other sources of technological contribution. We asked about the importance of scientific research in general and university-based research in particular. We also asked about the extent to which interindustry spillovers are an important source of technological opportunity. Part 4 asked some broad questions about the pace and character of technological advance.¹⁷ This paper analyzes responses to the questions in parts 1 and 2.¹⁸

SAMPLE CONSTRUCTION

As a sampling frame, we used the lines of business defined by the Federal Trade Commission. In the manufacturing sector, these chiefly correspond to four-digit SIC industries, although some are defined as groups of four-digit or even three-digit industries. The FTC lines provide the most disaggregated level at which data on R&D expenditures are available. An additional consideration was that F. M. Scherer's technology flow matrix, which classifies patents by industry of origin and industry of use, was also constructed at this level of aggregation.¹⁹

Ultimately, we received responses from 650 individuals representing 130 lines of business, with ten or more responses from eighteen industries and five to nine from twenty-seven industries. The sample was reasonably representative of firms performing R&D, though the exclusion of those without publicly traded securities undoubtedly means that small start-up ventures, important sources of innovation, were underrepresented. The number

WE USED THE LINES OF BUSINESS DEFINED BY THE FEDERAL TRADE COMMISSION...AN ADDITIONAL CONSIDERATION WAS THAT F. M. SCHERER'S TECHNOLOGY FLOW MATRIX, WHICH CLASSIFIES PATENTS BY INDUSTRY OF ORIGIN AND INDUSTRY OF USE, WAS ALSO CONSTRUCTED AT THIS LEVEL OF AGGREGATION.¹⁹

of respondents in a line of business was positively correlated with the line's R&D spending, sales volume, and R&D intensity. The number of respondents did not increase in strict proportion to the level of industry R&D or sales, but the rate of response within a line of business was not significantly correlated with industry R&D spending, sales, or R&D intensity. The Appendix presents further details of sample construction.

METHODOLOGICAL ISSUES

Given our interest in identifying differences in the appropriability of R&D, it is reassuring that analysis of variance confirmed the presence of significant interindustry variation in the responses to most questionnaire items.²⁰ There was, however, also substantial intraindustry variation in the responses.

There are several potential sources of intraindustry heterogeneity in the responses to any given question. First, the lines of business as defined by the FTC may be objectively heterogeneous in their products and technologies. For example, if two firms classified as manufacturers of industrial inorganic chemicals produce different products using different technologies, they might differ markedly in their perception of the effectiveness of patents or the time required for imitation in their "industry." To eliminate this source of heterogeneity, we asked respondents to identify two major innovations—a process and a product within their industries during the past ten to fifteen years. For most industries with ten or more respondents, more than half the respondents agreed on at least one such innovation. We thus believe it unlikely that overly aggregated industry definition was major source of intraindustry heterogeneity.²¹

A respondent's perception of the central tendencies within an industry may also be affected by his firm's policies or strategies. Respondents in the same line of business may thus have different perceptions of the common technological environment that they were asked to characterize. A two-way analysis of variance of the responses on the effectiveness of patents, for example, revealed that both firm and industry effects are statistically significant. A representative multi-industry firm, however, tends to be involved in technologically related industries, and thus what appear to be effects attributable to the firm in the data may simply reflect the correlation in responses from related industries.

The third, and probably most important, source of intraindustry heterogeneity is the inherently subjective nature of the semantic scales used in the survey. Most answers were reported on a seven-point Likert scale. The effectiveness of patents in preventing duplication was, for instance, evaluated on a scale ranging from “not at all effective” to “very effective.” There is no natural or objective anchor for such evaluative ratings. Individuals may perceive the same environment but simply use the scale differently. Some might systematically favor high scores; others might concentrate responses in the center of the scale; still others might frequently use extreme values.

The numerous techniques available to control for differences among respondents in means and variances generally require abandoning one or more dimensions along which the data might be informative. For example, we were interested in interindustry comparisons of answers to a single question; controlling for fixed effects among respondents would vitiate such comparisons, since we expected a respondent’s mean score over all questions to depend on his industry. Standardizing the variance of each respondent’s answers raised similar problems: the distribution of “correct” responses was unknown and it almost certainly differed systematically among industries. Rather than impose an arbitrary standardization, therefore, we examined the results for each group of questions using a variety of techniques and perspectives to assess the robustness of our principal conclusions. There was undeniably much noise in the data, but several important signals were robust to alternative weightings of the observations, alternative partitions of the sample, and the use of alternative summary statistics.²²

We sidestepped one methodological difficulty by treating ratings along a seven-point semantic continuum as if they were interval data. The data were, of course, more properly to be regarded as ordinal. It would have been straightforward to treat them as ordinal if we had been interested only in interindustry comparisons of responses to a single question. We also sought, however, to make comparisons among questions (for example, are patents more or less effective than secrecy in protecting process innovations from duplication?), and we therefore treated the data as if they were interval.²³

One additional methodological concern was whether our level of industry aggregation was appropriate for the problems being studied. The FTC line-of-business level was chosen to facilitate merging the data with disaggregated R&D data and Scherer’s classification of patents by industries of origin and use. Our analysis indicated, however, that most of the interesting interindustry distinctions among the 130 lines defined at the FTC level were robust to an aggregation of the data into the 25 industry groups used by the National Science

THE THIRD, AND PROBABLY MOST IMPORTANT, SOURCE OF INTRAINDUSTRY HETEROGENEITY IS THE INHERENTLY SUBJECTIVE NATURE OF THE SEMANTIC SCALES USED IN THE SURVEY. MOST ANSWERS WERE REPORTED ON A SEVEN-POINT LIKERT SCALE. THE EFFECTIVENESS OF PATENTS IN PREVENTING DUPLICATION WAS, FOR INSTANCE, EVALUATED ON A SCALE RANGING FROM “NOT AT ALL EFFECTIVE” TO “VERY EFFECTIVE.” THERE IS NO NATURAL OR OBJECTIVE ANCHOR FOR SUCH EVALUATIVE RATINGS. INDIVIDUALS MAY PERCEIVE THE SAME ENVIRONMENT BUT SIMPLY USE THE SCALE DIFFERENTLY. SOME MIGHT SYSTEMATICALLY FAVOR HIGH SCORES; OTHERS MIGHT CONCENTRATE RESPONSES IN THE CENTER OF THE SCALE; STILL OTHERS MIGHT FREQUENTLY USE EXTREME VALUES.

Foundation in its annual survey of R&D spending and employment patterns.

PATENTS AND OTHER MEANS OF APPROPRIATION

OUR ANALYSIS INDICATED, HOWEVER, THAT MOST OF THE INTERESTING INTERINDUSTRY DISTINCTIONS AMONG THE 130 LINES DEFINED AT THE FTC LEVEL WERE ROBUST TO AN AGGREGATION OF THE DATA INTO THE 25 INDUSTRY GROUPS USED BY THE NATIONAL SCIENCE FOUNDATION IN ITS ANNUAL SURVEY OF R&D SPENDING AND EMPLOYMENT PATTERNS.

Table 1 shows the pattern of responses, based on a seven-point scale, to questions on the effectiveness of alternative means of capturing and protecting the competitive advantages of new or improved processes and products. The first two columns report the mean response for the entire sample of 650 respondents to each question, as well as the standard error of each estimated mean. These statistics, of course, give equal weight to each respondent and consequently weight each industry in proportion to its number of respondents. The overall pattern across questions, however, is robust to the use of alternative summary statistics,

such as the mean of industry means or the median of industry means. This is apparent in columns 3 and 4, which summarize the distribution of industry mean responses to each question. Each pair of numbers represents the range of industry means from the upper bound of the lowest quintile to the lower bound of the highest quintile of industries: 20 percent of the 130 industries had mean responses at or below the bottom of the range indicated for each question, and 20 percent had mean responses at or above the top of the range. Mean responses for the remaining 60 percent (or 78 industries) fell within the reported range.

Table 1. Effectiveness of Alternative Means of Protecting the Competitive Advantages of New or Improved Processes and Products^a

<i>Method of appropriation</i>	<i>Overall sample means</i>		<i>Distribution of industry means^b</i>	
	<i>Processes</i>	<i>Products</i>	<i>Processes</i>	<i>Products</i>
Patents to prevent duplication	3.52 (0.06)	4.33 (0.07)	2.6–4.0 ^c	3.0–5.0 ^c
Patents to secure royalty income	3.31 (0.06)	3.75 (0.07)	2.3–4.0 ^c	2.7–4.8 ^c
Secrecy	4.31 (0.07)	3.57 (0.06)	3.3–5.0	2.7–4.1
Lead time	5.11 (0.05)	5.41 (0.05)	4.3–5.9 ^c	4.8–6.0 ^c
Moving quickly down the learning curve	5.02 (0.05)	5.09 (0.05)	4.5–5.7	4.4–5.8
Sales or service efforts	4.55 (0.07)	5.59 (0.05)	3.7–5.5	5.0–6.1

Source: Authors' calculations.

a. Range: 1 = not at all effective; 7 = very effective. Standard errors in parentheses.

b. From the upper bound of the lowest quintile of industries to the lower bound of the highest quintile.

c. Differences in means significant at the .01 level.

The picture is striking. For new processes (columns 1 and 3), patents were generally rated the least effective of the mechanisms of appropriation: only 20 percent of the lines of business surveyed rated process patent effectiveness in excess of 4.0. Eighty percent scored the effectiveness of lead time and learning curve advantages on new processes in excess of 4.3. Secrecy, though not considered as effective as lead-time and learning advantages,

was still considered more effective than patents in protecting processes.

Patents for products were typically considered more effective than those for processes, and secrecy was considered less effective in protecting products than processes. Generally, lead-time, learning curves, and sales or service efforts were regarded as substantially more effective than patents in protecting products. Eighty percent of the sample businesses rated the effectiveness of sales and service efforts above 5.0, but only 20 percent considered product patents this effective.²⁴

The tendency to regard secrecy as more effective than process patents but less effective than product patents probably reflects the greater ease and desirability of maintaining secrecy about process technology.

Firms may sometimes refrain from patenting processes to avoid disclosing either the fact or the details of an innovation.²⁵ But firms have every incentive to advertise the advantages of new or improved products and to get them into the hands of customers, thereby facilitating direct observation of the product and the technology it embodies. Maintaining secrecy about product innovations is thus likely to be both difficult and undesirable.

Respondents also tended to regard patents to prevent duplication as more effective than patents to secure royalty income. This finding was consistent with the view that licensing arrangements are beset with transactional difficulties.

Only 3 of 130 lines of business rated process patents higher than five on a seven-point scale of effectiveness in preventing duplication. Two of these were concrete and primary copper; the other had only a single respondent.²⁶ Only 5 of 130 industries rated product patents to prevent duplication higher than six points. Two of these were singletons; the other three were drugs, pesticides, and industrial organic chemicals. Twenty other

THE PICTURE IS STRIKING. FOR NEW PROCESSES (COLUMNS 1 AND 3), PATENTS WERE GENERALLY RATED THE LEAST EFFECTIVE OF THE MECHANISMS OF APPROPRIATION: ONLY 20 PERCENT OF THE LINES OF BUSINESS SURVEYED RATED PROCESS PATENT EFFECTIVENESS IN EXCESS OF 4.0.

lines rated product patents between five and six. Of those with more than two responses, almost all fell neatly into chemical products (including inorganic chemicals, plastic materials, synthetic fibers, synthetic rubber, and glass) or relatively uncomplicated mechanical equipment (air and gas compressors, power-driven hand tools, and oilfield machinery). The only anomalies were roasted coffee and products of steel rolling and finishing mills.

Table 2 shows additional industry-level detail—the mean rating given for the effectiveness of patents in preventing duplication in eighteen industries with ten or more respondents. These industries tend to be much

TABLE 2 SHOWS ADDITIONAL INDUSTRY-LEVEL DETAIL—THE MEAN RATING GIVEN FOR THE EFFECTIVENESS OF PATENTS IN PREVENTING DUPLICATION IN EIGHTEEN INDUSTRIES WITH TEN OR MORE RESPONDENTS. THESE INDUSTRIES TEND TO BE MUCH MORE RESEARCH-INTENSIVE THAN THE SAMPLE AVERAGE, YET THE PATTERN OF INTERINDUSTRY VARIATION WAS SIMILAR TO THAT IN THE FULL SAMPLE.

more research-intensive than the sample average, yet the pattern of interindustry variation was similar to that in the full sample. Except for petroleum refining, product patents were considered more effective than process patents. Only four chemical industries (drugs, plastic materials, inorganic chemicals, and organic chemicals)

Table 2. Effectiveness of Process and Product Patents in Industries with Ten or More Survey Responses

<i>Industry</i>	<i>Process patents</i>		<i>Product patents</i>	
	<i>Mean</i>	<i>Standard error</i>	<i>Mean</i>	<i>Standard error</i>
Pulp, paper, and paperboard	2.6	0.3	3.3	0.4
Cosmetics	2.9	0.3	4.1	0.4
Inorganic chemicals	4.6	0.4	5.2	0.3
Organic chemicals	4.1	0.3	6.1	0.2
Drugs	4.9	0.3	6.5	0.1
Plastic materials	4.6	0.3	5.4	0.3
Plastic products	3.2	0.3	4.9	0.3
Petroleum refining	4.9	0.4	4.3	0.4
Steel mill products	3.5	0.7	5.1	0.6
Pumps and pumping equipment	3.2	0.4	4.4	0.5
Motors, generators, and controls	2.7	0.3	3.5	0.5
Computers	3.3	0.4	3.4	0.4
Communications equipment	3.1	0.3	3.6	0.3
Semiconductors	3.2	0.4	4.5	0.4
Motor vehicle parts	3.7	0.4	4.5	0.4
Aircraft and parts	3.1	0.5	3.8	0.4
Measuring devices	3.6	0.3	3.9	0.3
Medical instruments	3.2	0.4	4.7	0.4
Full sample	3.5	0.06	4.3	0.07

Source: Authors' calculations. Mean score on a scale of 1 to 7.

and petroleum refining rated process patent effectiveness higher than four on a seven point scale, and only these four chemical industries and steel mills rated product patents higher than five.²⁷

The data on these eighteen most heavily sampled industries help to establish the robustness of our conclusion about the limited effectiveness of patents as a means of appropriation. In none did a majority of respondents rate patents-either to prevent duplication or to secure royalty income-as more effective than the most highly rated of the other four means of appropriating returns from new processes, although in drugs and petroleum

IN NONE DID A MAJORITY OF RESPONDENTS RATE PATENTS-EITHER TO PREVENT DUPLICATION OR TO SECURE ROYALTY INCOME-AS MORE EFFECTIVE THAN THE MOST HIGHLY RATED OF THE OTHER FOUR MEANS OF APPROPRIATING RETURNS FROM NEW PROCESSES, ALTHOUGH IN DRUGS AND PETROLEUM REFINING A MAJORITY REGARDED PROCESS PATENTS AS AT LEAST THE EQUAL OF THE MOST EFFECTIVE ALTERNATIVE MECHANISM OF APPROPRIATION.

refining a majority regarded process patents as at least the equal of the most effective alternative mechanism of appropriation. In only one industry, drugs were product patents regarded by a majority of respondents as strictly more effective than other means of appropriation.²⁸ In three others-organic chemicals, plastic materials, steel mill products most respondents rated patents as no less effective than the best alternative.

The exclusion from our sample of firms that offered no publicly traded securities may have biased our findings. For small, start-up ventures, patents may be a relatively effective means of appropriating R&D returns, in part because some other means, such as investment in complementary sales and service efforts,

may not be feasible. The patents held by a small, technologically oriented firm may be its most marketable asset. Although our respondents were asked to describe the typical experience of firms in their industries, they may well have overlooked aspects of appropriability that are particularly relevant for new firms.

The most probable explanation for the robust finding that patents are particularly effective in chemical industries is that comparatively clear standards can be applied to assess a chemical patent's validity and to defend against infringement. The uniqueness of a specific molecule is more easily demonstrated than the novelty of, for example, a new component of a complex electrical or mechanical system. Similarly, it is easy to determine whether an allegedly infringing molecule is physically identical to a patented molecule; it is more difficult to determine whether comparable components of two complex systems "do the same work in substantially the same way." To the extent that very simple mechanical inventions approximate molecules in their discreteness and easy differentiability, it is understandable that industries producing such machinery rank just after chemical industries in the perceived effectiveness of patent protection.

The perceived ineffectiveness of patents in most industries raises the question of why firms use them. Further work is needed here, but we offer some speculations informed by the comments of our pretest subjects and by several survey respondents at a conference we held to report preliminary findings. These executives identified two motives for patenting that have little connection with appropriating returns from investment. One is to measure the performance of R&D employees, which is a significant problem because these workers are typically engaged in team production. Legal standards for identifying inventors on a patent application are, however, reasonably rigorous. The second motive is to gain access to certain foreign markets. Some developing countries require, as a condition of entry, that U.S. firms

SOME DEVELOPING COUNTRIES REQUIRE, AS A CONDITION OF ENTRY, THAT U.S. FIRMS LICENSE TECHNOLOGY TO A HOST-COUNTRY FIRM, AND SOME PATENTS ARE FILED PRIMARILY TO PERMIT SUCH LICENSING.²⁹

THE PERCEIVED INEFFECTIVENESS OF PATENTS IN MOST INDUSTRIES RAISES THE QUESTION OF WHY FIRMS USE THEM. FURTHER WORK IS NEEDED HERE, BUT WE OFFER SOME SPECULATIONS INFORMED BY THE COMMENTS OF OUR PRETEST SUBJECTS AND BY SEVERAL SURVEY RESPONDENTS AT A CONFERENCE WE HELD TO REPORT PRELIMINARY FINDINGS. THESE EXECUTIVES IDENTIFIED TWO MOTIVES FOR PATENTING THAT HAVE LITTLE CONNECTION WITH APPROPRIATING RETURNS FROM INVESTMENT. ONE IS TO MEASURE THE PERFORMANCE OF R&D EMPLOYEES, WHICH IS A SIGNIFICANT PROBLEM BECAUSE THESE WORKERS ARE TYPICALLY ENGAGED IN

license technology to a host-country firm, and some patents are filed primarily to permit such licensing.²⁹

CONDITIONS AFFECTING APPROPRIABILITY

Thus far we have focused on the overall strength of various mechanisms of appropriation and on interindustry

variations in the effectiveness of patents. The patterns of covariation in the responses, however, suggested that interindustry differences in conditions affecting appropriability might be summarized by a limited number of factors. Moreover, the clear indications that patents are effective in only a few industries suggested that it might be fruitful to classify industries into clusters distinguishable by a primary means of appropriation and

YET ANOTHER MOTIVE DISCUSSED IN THE LITERATURE IS TO GAIN STRATEGIC ADVANTAGE IN NEGOTIATION. IN THE SEMICONDUCTOR INDUSTRY, FOR INSTANCE, THE CUMULATIVE NATURE OF THE TECHNOLOGY MAKES IT DIFFICULT TO PARTICIPATE LEGALLY WITHOUT ACCESS TO THE PATENTS OF NUMEROUS FIRMS. IN CONSEQUENCE, THERE IS WIDESPREAD CROSS-LICENSING. ESTABLISHED FIRMS, HOWEVER, RARELY LICENSE A NEW ENTRANT UNTIL IT HAS ESTABLISHED A SIGNIFICANT POSITION IN THE MARKET.

perhaps by the overall ease of appropriating returns. Such clusters could prove useful in examining links between appropriability conditions and measures of R&D, innovation, and productivity growth.

Correlations among responses to questions on the effectiveness of alternative means of appropriation revealed some interesting patterns.³⁰ When patents effectively prevent competitors from duplicating processes and products, they tend also to be effective in securing royalty income. But neither form of effectiveness was strongly correlated with the effectiveness of other means of appropriation. For processes, there was a strong connection among three other mechanisms: lead time, learning curve advantages, and secrecy. For products, superior sales and service efforts were strongly linked to lead time and learning advantages, though not to secrecy.

The correlations suggested that the mechanisms of appropriation might be reduced to two dimensions: one associated with the use of patents, the other related to secrecy, lead time, and learning curve advantages. For product innovations, sales and service efforts may be involved in the second of these dimensions. We investigated this possibility by reducing the data to principal components and employing a variety of factor-analytic techniques. Principal factor analysis and several methods of rotation did little to alter the picture presented by the principal components, which are shown in table 3.³¹

The first two columns of the table show the weights associated with the first two principal components when the six questions relating to process appropriability are analyzed separately from the six questions relating to product appropriability. The next two columns report the results of a principal components analysis on the entire set of twelve questions. With both approaches, the first principal component gives near-zero weight to the two patent-related methods of appropriation and heavy weight to the other mechanisms. The weighting is reversed for the second principal component. Thus the first two principal components (and, in the factor analysis, the first two factors) are readily interpreted, respectively, as nonpatent - and patent-related dimensions of appropriability. Despite this clear interpretation, the data do not reduce very satisfactorily to just two dimensions. As table 3 indicates, when the process and product questions are analyzed separately, the first two components explain only 60 percent of the variance in the responses to six questions, and when the two sets of questions are combined, two components explain only 50 percent of the variance.

THE CORRELATIONS SUGGESTED THAT THE MECHANISMS OF APPROPRIATION MIGHT BE REDUCED TO TWO DIMENSIONS: ONE ASSOCIATED WITH THE USE OF PATENTS, THE OTHER RELATED TO SECRECY, LEAD TIME, AND LEARNING CURVE ADVANTAGES.

Table 3. Principal Components Analysis of Methods of Appropriation

<i>Method of appropriation</i>	<i>Processes and products separately</i>		<i>Processes and products together</i>	
	<i>Coefficients of 1st principal component</i>	<i>Coefficients of 2d principal component</i>	<i>Coefficients of 1st principal component</i>	<i>Coefficients of 2d principal component</i>
	New Processes			
Patents to prevent duplication	.04	.86	.01	.73
Patents to secure royalties	.12	.86	.08	.78
Secrecy	.59	-.12	.54	.04
Lead time	.84	-.09	.79	-.04
Moving down the learning curve	.84	-.05	.80	-.04
Sales and service efforts	.51	.11	.45	-.06
Cumulative variance explained	.34	.59	n.a.	n.a.
New Products				
Patents to prevent duplication	.06	.87	.06	.73
Patents to secure royalties	.06	.87	.07	.80
Secrecy	.51	.01	.51	.06
Lead time	.84	.00	.79	-.03
Moving down the learning curve	.84	-.07	.82	-.04
Sales and service efforts	.69	-.09	.62	-.11
Cumulative variance explained	.36	.61	.31	.50

Source: Authors' calculations.

Our interpretation that the means of appropriation can be grouped into patent and nonpatent mechanisms was nonetheless reinforced by a cluster analysis that classified industries according to mean responses to the relevant questions. The best clustering results were achieved by dividing the industries into three groups, as shown in table 4. Industries assigned to cluster 1 tended to have relatively low scores for all mechanisms of appropriation. Sales and service effort was the most highly rated mechanism and was, in fact, regarded as reasonably effective in capturing returns from new products. Industries in cluster 2 rated lead-time and learning curves as relatively effective, but not patents. Secrecy was important in appropriating process returns, and sales and service efforts complemented lead-time and learning advantages for products. Only for cluster 3 were product and process patents deemed effective, but still the effectiveness of lead-time and learning was no lower than for the industries in cluster 2. Those few industries in which patents were rated as more effective than other mechanisms were all in the third cluster.

Table 4. Cluster Analysis of Mechanisms of Appropriation

<i>Method of appropriation</i>	<i>Cluster</i>		
	<i>1</i>	<i>2</i>	<i>3</i>
New Processes			
Number of industries	38	67	25
Mean score			
Patents to prevent duplication	3.1	3.0	4.7
Patents to secure royalties	2.9	2.9	4.8
Secrecy	2.8	4.6	4.7
Lead time	4.2	5.4	5.6
Learning curves	4.3	5.3	5.1
Superior sales or service	4.7	4.5	4.9
New Products			
Number of industries	20	68	42
Mean score			
Patents to prevent duplication	3.1	3.8	5.3
Patents to secure royalties	3.2	3.1	5.0
Secrecy	2.6	3.5	4.0
Lead time	4.0	5.6	5.7
Learning curves	4.2	5.3	5.2
Superior sales or service	5.2	5.7	5.6

Source: Authors' calculations.

OUR INTERPRETATION THAT THE MEANS OF APPROPRIATION CAN BE GROUPED INTO PATENT AND NONPATENT MECHANISMS WAS NONETHELESS REINFORCED BY A CLUSTER ANALYSIS THAT CLASSIFIED INDUSTRIES ACCORDING TO MEAN RESPONSES TO THE RELEVANT QUESTIONS.

The cluster analysis suggested that there was a group of industries in which no appropriation mechanism was particularly effective. As an alternative approach to identifying settings with low appropriability, we considered the maximum score an industry assigned to any of the six mechanisms on the questionnaire. Only 11 of the 130 failed to rate at least one means of appropriating returns from product innovation higher than five on the seven-point scale. The industries in this group with more than two responses were all drawn from the food products and metal working sectors: milk, meat

products, iron and steel foundries, boiler shops, and screw machine products (nuts, bolts, and screws). Many more industries (34 of 130) rated no means of appropriating process returns higher than five. This group contained all the industries (except milk) that ranked low on product appropriability but was otherwise a diverse lot. The heaviest concentration was in fabricated metals and machinery. But several chemical industries were also represented, including the three industries in which product patents were viewed as most effective—organic chemicals, pesticides, and drugs.

The urge to find patterns in the data should not be carried too far. The associations among mechanisms of appropriation revealed by the correlation, principal components, and cluster analyses are suggestive, but there is substantial heterogeneity in the underlying data. As noted, the first two principal components, though readily interpretable, explained an unsatisfactory fraction of the overall variance. A similar lack of good fit characterized the cluster analyses of process and product appropriability. Despite the fairly clear interpretation that could be given to each cluster, the variance within the clusters was almost twice that between clusters.

LIMITATIONS ON EFFECTIVENESS OF PATENTS

To understand why patent protection might be weak in some industries, we asked respondents to rate the importance of possible limitations on patent effectiveness. Table 5 summarizes the responses. The ability of competitors to “invent around” both process and product patents were rated higher than five on a seven-point scale of importance by 60 percent of the responding industries. Only one other constraint—the lack of ready patentability for new processes—was rated this important by more than 20 percent. Limitations on patents were generally considered more severe for processes than for products, which was consistent with our finding that product patents tend to be more effective than process patents. In particular, the lack of patentability was more serious for processes than for products, and so was the disclosure of information through patent documents.³²

The responses concerning limits on patent effectiveness may illuminate and focus policy discussion.

TO UNDERSTAND WHY PATENT PROTECTION MIGHT BE WEAK IN SOME INDUSTRIES, WE ASKED RESPONDENTS TO RATE THE IMPORTANCE OF POSSIBLE LIMITATIONS ON PATENT EFFECTIVENESS...THE ABILITY OF COMPETITORS TO “INVENT AROUND” BOTH PROCESS AND PRODUCT PATENTS WERE RATED HIGHER THAN FIVE ON A SEVEN-POINT SCALE OF IMPORTANCE BY 60 PERCENT OF THE RESPONDING INDUSTRIES.

Table 5. Limitations on Effectiveness of Patents for New or Improved Processes and Products^a

Limitation	Overall sample means		Distribution of industry means ^b	
	Processes	Products	Processes	Products
New processes or products patentable	4.32 (0.07)	3.75 (0.07)	3.6–5.4 ^c	2.8–4.8
Patents unlikely to be valid if challenged	4.18 (0.06)	3.92 (0.07)	3.5–5.0 ^c	3.0–5.0 ^c
Firms do not enforce patents	4.29 (0.06)	3.84 (0.07)	3.5–5.0 ^c	3.0–4.8 ^c
Competitors legally “invent around” patents	5.49 (0.05)	5.09 (0.06)	4.9–6.0	4.4–5.9 ^d
Technology moving so fast that patents are irrelevant	3.40 (0.07)	3.34 (0.07)	2.0–4.3 ^d	2.0–4.0 ^d
Patent documents disclose too much information	4.19 (0.07)	3.65 (0.07)	3.2–5.0	2.8–4.5 ^c
Licensing required by court decisions	2.96 (0.06)	2.79 (0.06)	2.0–3.8	2.0–3.3
Firms participate in cross-licensing agreements with competitors	3.08 (0.06)	2.93 (0.06)	2.2–3.9 ^d	2.1–3.9 ^d

Source: Authors' calculations.

a. Range: 1 = not an important limitation; 7 = very important limitation. Standard errors in parentheses.

b. From the upper bound of the lowest quintile of industries to the lower bound of the highest quintile.

c. Interindustry differences significant at the .10 level.

d. Interindustry differences significant at the .01 level.

In recent years there has been considerable interest in making patent protection more effective. One initiative has been to make the legal requirements for a valid patent claim less stringent.³³ Another has been to vacate court decrees that compel licensing. Our data identified industries in which stringent requirements for patent validity or compulsory licensing were perceived as important limitations on the usefulness of patents in appropriating returns.

Respondents from twenty-two lines of business, mostly in the food processing and fabricated metals sectors, considered the likely inability to withstand challenges *to* validity as significantly limiting the effectiveness of process patents (scoring the importance higher than five on a seven-point scale); for fourteen of these industries the mean response was six or higher on the scale. This group and the nineteen industries citing invalidity as a constraint on the effectiveness of product patents (again assigning a score higher than five) overlapped considerably. Further investigation would be required to determine just why these two sectors appear to have difficulty establishing valid claims. Perhaps because they are mature industries, opportunities may be limited or novelty may be difficult to achieve or simply difficult to prove.

Compulsory licensing was rarely judged a significant limit on the effectiveness of patents. Only one industry with one respondent rated this constraint higher than five on the scale for products, and only six cited compulsory licensing of process patents as of comparable importance. Two of these industries were

THE RESPONSES CONCERNING LIMITS ON PATENT EFFECTIVENESS MAY ILLUMINATE AND FOCUS POLICY DISCUSSION. IN RECENT YEARS THERE HAS BEEN CONSIDERABLE INTEREST IN MAKING PATENT PROTECTION MORE EFFECTIVE. ONE INITIATIVE HAS BEEN TO MAKE THE LEGAL REQUIREMENTS FOR A VALID PATENT CLAIM LESS STRINGENT.³³ ANOTHER HAS BEEN TO VACATE COURT DECREES THAT COMPEL LICENSING. OUR DATA IDENTIFIED INDUSTRIES IN WHICH STRINGENT REQUIREMENTS FOR PATENT VALIDITY OR COMPULSORY LICENSING WERE PERCEIVED AS IMPORTANT LIMITATIONS ON THE USEFULNESS OF PATENTS IN APPROPRIATING RETURNS.

COMPULSORY LICENSING WAS RARELY JUDGED A SIGNIFICANT LIMIT ON THE EFFECTIVENESS OF PATENTS. ONLY ONE INDUSTRY WITH ONE RESPONDENT RATED THIS CONSTRAINT HIGHER THAN FIVE ON THE SCALE FOR PRODUCTS, AND ONLY SIX CITED COMPULSORY LICENSING OF PROCESS PATENTS AS OF COMPARABLE IMPORTANCE.... COMPULSORY LICENSING DECREES WERE THUS PERCEIVED AS IMPORTANT IN ONLY A SMALL SUBSET OF THE INDUSTRIES THAT F. M. SCHERER INDICATED WERE SUBJECT TO SUCH DECREES.³⁴ THE OVERALL LACK OF IMPACT FROM COMPULSORY LICENSING REQUIREMENTS WAS CONSISTENT WITH SCHERER'S FINDING THAT THEY DID NOT DISCOURAGE R&D SPENDING.

not singletons-metal containers and electron tubes. Compulsory licensing decrees were thus perceived as important in only a small subset of the industries that F. M. Scherer indicated were subject to such decrees.³⁴ The overall lack of impact from compulsory licensing requirements was consistent with Scherer's finding that they did not discourage R&D spending.

The choice between obtaining a patent and maintaining secrecy may be influenced by the extent to which the disclosures made in the patent document facilitate inventing around the patent. Our data provided some support for this theory. The effectiveness of secrecy was positively correlated with the extent to which disclosures limited the effectiveness of patents. The link was stronger for product patents than for process patents. But patent disclosures represented a substantial limitation on the effectiveness of product patents for only 4 of the 130 industries (scoring as high as six on the scale), and only 16 regarded process disclosures as comparably important. In only one line of business of those with five or more respondents-metal cutting machine tools-did disclosures constrain so substantially the effectiveness of both process and product patents.

Table 6. Effectiveness of Alternative Methods of Learning about New Processes and Products^a

<i>Method of learning</i>	<i>Overall sample means</i>		<i>Distribution of industry means^b</i>	
	<i>Processes</i>	<i>Products</i>	<i>Processes</i>	<i>Products</i>
Licensing technology	4.58 (0.07)	4.62 (0.07)	3.4–5.6 ^c	3.5–5.5 ^c
Patent disclosures	3.88 (0.05)	4.01 (0.06)	3.0–4.6 ^c	3.0–4.8 ^c
Publications or technical meetings	4.07 (0.05)	4.07 (0.05)	3.4–4.7	3.3–4.6 ^d
Conversations with employees of innovating firm	3.64 (0.06)	3.64 (0.06)	2.9–4.7 ^d	2.9–4.5 ^d
Hiring R&D employees from innovating firm	4.02 (0.07)	4.08 (0.07)	2.7–5.0 ^c	2.8–5.0 ^c
Reverse engineering of product	4.07 (0.07)	4.83 (0.06)	3.0–5.0 ^c	4.0–5.7 ^d
Independent R&D	4.76 (0.06)	5.00 (0.05)	4.0–5.5	4.4–5.6 ^c

Source: Authors' calculations.

a. Range: 1 = not at all effective; 7 = very effective. Standard errors in parentheses.

b. From the upper bound of the lowest quintile of industries to the lower bound of the highest quintile.

c. Interindustry differences in means significant at the .01 level.

d. Interindustry differences in means significant at the .05 level.

CHANNELS OF INFORMATION SPILLOVER

To the extent that a rival can learn easily about an innovator's technology, the incentive to invest in R&D is attenuated. But to the extent that learning is easy, wasteful duplication or near duplication of R&D effort by rival firms may be avoided. Also, knowledge of an innovator's new technology may complement rival R&D effort by enhancing its productivity. Richard Nelson and Sidney Winter, Michael Spence, and Richard Levin and Peter C. Reiss have developed models that begin to disentangle these offsetting effects, called by Spence the incentive and efficiency effects of interfirm spillovers.³⁵ A sharper characterization of interindustry differences in the nature and strength of the mechanisms by which firms learn about their competitors' technology should advance these modeling efforts.

THE CHOICE BETWEEN OBTAINING A PATENT AND MAINTAINING SECRECY MAY BE INFLUENCED BY THE EXTENT TO WHICH THE DISCLOSURES MADE IN THE PATENT DOCUMENT FACILITATE INVENTING AROUND THE PATENT. OUR DATA PROVIDED SOME SUPPORT FOR THIS THEORY. THE EFFECTIVENESS OF SECRECY WAS POSITIVELY CORRELATED WITH THE EXTENT TO WHICH DISCLOSURES LIMITED THE EFFECTIVENESS OF PATENTS.

Table 6 summarizes the responses to questions about the effectiveness of alternative ways of learning. There is little difference between the pattern of responses for processes and for products, except that, as one would expect, reverse engineering is markedly more effective in yielding information about product technology. On average, independent R&D was rated as the most effective means of learning about rival technology¹³⁶ this

Table 7. Cluster Analysis of Channels of Learning

<i>Learning mechanism</i>	<i>Cluster</i>		
	<i>1</i>	<i>2</i>	<i>3</i>
New Processes			
Number of industries	68	43	19
Mean score			
Licensing technology	5.0	4.3	2.5
Patent disclosures	4.0	4.0	3.2
Publications or technical meetings	3.8	4.6	3.9
Conversations with employees of innovating firm	3.2	4.8	3.0
Hiring R&D employees from innovating firm	3.7	5.1	2.4
Reverse engineering of product	3.8	4.6	4.0
Independent R&D	5.0	4.6	4.3
New products			
Number of industries	68	10	52
Mean score			
Licensing technology	4.7	2.5	4.5
Patent disclosures	3.9	2.9	4.3
Publications or technical meetings	3.7	5.1	4.3
Conversations with employees of innovating firm	3.0	4.6	4.5
Hiring R&D employees from innovating firm	3.2	4.4	4.9
Reverse engineering of product	4.7	3.0	5.2
Independent R&D	5.1	3.7	5.0

Source: Authors' calculations.

may appear to be wasteful duplication, but it need not be. One pretest subject said that R&D effort devoted to determining what a competitor has done may have strong complementarities with a firm's own research program in areas not directly imitative of the innovating competitor. Licensing was also rated, on average, an

important way of gaining access to a rival's new technology.

The correlations between individual and industry mean responses show that mechanisms relying on interpersonal communication (publications and technical meetings, informal conversations, and hiring away employees) are strongly inter-correlated. Learning through licensing technology is uncorrelated with nearly all other learning mechanisms except disclosure through patent documents. There are two possible interpretations of this last connection. Potential licensees may learn about the opportunity to license through patent documents, or the documents may prove useful in employing new technology once it is licensed. We cannot tell whether the “announcement” effect or the “complementary information” effect of disclosures predominates.³⁷

THE PATTERN OF CORRELATION SUGGESTED THAT THERE MIGHT BE THREE OR FOUR CLUSTERS OF INDUSTRIES, DISTINGUISHED IN TURN BY AN EMPHASIS ON LEARNING THROUGH LICENSING, INTERPERSONAL CHANNELS, AND REVERSE ENGINEERING OR INDEPENDENT R&D, OR BOTH. THE RESULTS OBTAINED FROM CLUSTER ANALYSIS WERE NOT ENTIRELY SATISFACTORY.³⁸

The pattern of correlation suggested that there might be three or four clusters of industries, distinguished in turn by an emphasis on learning through licensing, interpersonal channels, and reverse engineering or independent R&D, or both. The results obtained from cluster analysis were not entirely satisfactory.³⁸ Nonetheless, table 7 presents the results of grouping the lines of business into three clusters on the basis of responses to the questions on channels of spillover.

For both new processes and products, the largest group of industries typically relied on licensing and independent R&D to learn about competitive technology. Interpersonal channels were relatively unimportant, and reverse engineering was important for products. For both processes and products, there was a second cluster of industries in which interpersonal channels of spillover were most important. In the case of learning about new products, only ten industries were classified in this cluster, and in the case of learning about new processes, other channels independent R&D and reverse engineering-were nearly as valuable. For processes, a third cluster appeared to find all mechanisms of learning relatively unproductive. For products this group found all mechanisms moderately effective.

For both new processes and products, the largest group of industries typically relied on licensing and independent R&D to learn about competitive technology. Interpersonal channels were relatively unimportant, and reverse engineering was important for products. For both processes and products, there was a second cluster of industries in which interpersonal channels of spillover were most important. In the case of learning about new products, only ten industries were classified in this cluster, and in the case of learning about new processes, other channels independent R&D and reverse engineering-were nearly as valuable. For processes, a third cluster appeared to find all mechanisms of learning relatively unproductive. For products this group found all mechanisms moderately effective.

COST AND TIME REQUIRED FOR IMITATION

As part of our investigation we asked respondents to estimate typical costs and time required to duplicate several categories of innovations if a competitor developed them. For each category, respondents were asked to identify (within a range) the cost of duplication as a percentage of the innovator's R&D cost. Intervals measured in months or years were used to classify the time required. In light of evidence that there is a time-cost trade-off in certain industries, we asked respondents to estimate the cost and time required “to have a significant impact on the market.”³⁹

Tables 8 and 9 show frequency distributions of industry median responses.⁴⁰ The dispersion of industry medians suggests substantial variations among industries in both the cost and time required to duplicate all categories of innovation. If, however, individual responses to the questions on cost are coded on a six-point interval scale, there is sufficient intraindustry variation to render interindustry differences insignificant at the

Table 8. Cost of Duplicating an Innovation as a Percentage of Innovator's R&D Cost, Frequency Distribution of Median Responses

Type of innovation	Less than 25 percent	26 to 50 percent	51 to 75 percent	76 to 100 percent	More than 100 percent	Timely duplication not possible
New process						
Major patented new process	1	5	19	66	26	10
Major unpatented new process	5	10	55	49	6	2
Typical patented new process	2	15	61	41	6	2
Typical unpatented new process	8	43	58	14	4	0
New product						
Major patented new product	1	4	17	63	30	12
Major unpatented new product	5	13	58	40	7	4
Typical patented new product	2	18	64	32	9	2
Typical unpatented new product	9	58	40	15	5	0

Source: Survey of 127 lines of business.

0.01 levels. Interindustry differences in the time required for duplication are, by contrast, significant at the 0.01 level in every instance except the time required to duplicate a typical patented new process.

Several conclusions are apparent. First, duplicating major innovations tends to cost more and take longer than duplicating typical innovations. (In a sense, this confirms that respondents correctly interpreted the distinction between typical and major innovations.) Second, for a given category of innovation, the cost and time required to duplicate are distributed very similarly for products and processes. Products tend to be slightly cheaper and quicker to duplicate than processes, though this generalization does not hold for major patented innovations. Finally, patents tend to raise imitation costs and time for each category of innovation. These increases can be regarded as alternative indicators of the relative effectiveness of patents in different industries.

To explore this point further, we coded the individual responses to the imitation costs and time questions on a six-point interval scale, calculated the individual and industry mean increases in costs and time associated with the presence of patents, and correlated these, respectively, with individual and industry mean responses to our questions on the effectiveness of patents in preventing duplication. For each category

SEVERAL CONCLUSIONS ARE APPARENT. FIRST, DUPLICATING MAJOR INNOVATIONS TENDS TO COST MORE AND TAKE LONGER THAN DUPLICATING TYPICAL INNOVATIONS. (IN A SENSE, THIS CONFIRMS THAT RESPONDENTS CORRECTLY INTERPRETED THE DISTINCTION BETWEEN TYPICAL AND MAJOR INNOVATIONS.) SECOND, FOR A GIVEN CATEGORY OF INNOVATION, THE COST AND TIME REQUIRED TO DUPLICATE ARE DISTRIBUTED VERY SIMILARLY FOR PRODUCTS AND PROCESSES. PRODUCTS TEND TO BE SLIGHTLY CHEAPER AND QUICKER TO DUPLICATE THAN PROCESSES, THOUGH THIS GENERALIZATION DOES NOT HOLD FOR MAJOR PATENTED INNOVATIONS. FINALLY, PATENTS TEND TO RAISE IMITATION COSTS AND TIME FOR EACH CATEGORY OF INNOVATION.

**Table 9. time Required to Duplicate an Innovation,
Frequency Distribution of Median Responses**

Type of innovation	Less than 6 months	6 months to 1 year	1 to 3 years	3 to 5 years	More than 5 years	Timely duplication not possible
New process						
Major patented new process	0	4	72	37	9	7
Major unpatented new process	2	20	84	17	2	4
Typical patented new process	0	40	73	13	0	3
Typical unpatented new process	8	66	47	6	1	1
New product						
Major patented new product	2	6	64	40	8	9
Major unpatented new product	3	22	89	12	1	2
Typical patented new product	5	39	72	6	4	3
Typical unpatented new product	18	67	39	4	1	0

Source: Survey of 129 lines of business.

of innovation, the reported effectiveness of patents was positively correlated with the increase in duplication costs and time associated with patents, although the correlations tended to be stronger for products than for processes. We also found some evidence, at the level of the individual respondent, that patent effectiveness was associated with the absolute level of duplication costs for patented processes and products. We found a much stronger association, however, between reported patent effectiveness and the amount of time required to duplicate both patented process and product innovations.

These broad-brush patterns of association conceal some striking anomalies. For particular categories of innovation, at least two and as many as fourteen industries reported that patents actually reduced the costs or time required for duplication. A partial explanation is that a disproportionate number of these industries also reported that disclosure of information through patent documents was a significant limitation on patent effectiveness.

OUR RESULTS WERE CONSISTENT WITH THOSE OF MANSFIELD, SCHWARTZ, AND WAGNER. WE FOUND THAT PATENTS RAISE IMITATION COSTS BY 40 PERCENTAGE POINTS FOR BOTH MAJOR AND TYPICAL NEW DRUGS, BY 30 POINTS FOR MAJOR NEW CHEMICAL PRODUCTS, AND BY 25 POINTS FOR TYPICAL CHEMICAL PRODUCTS. IN ELECTRONICS, OUR RESULTS DIFFERED SOMEWHAT FOR SEMICONDUCTORS, COMPUTERS, AND COMMUNICATIONS EQUIPMENT, BUT THE RANGE WAS 7 TO 15 PERCENTAGE POINTS FOR MAJOR PRODUCTS AND 7 TO 10 FOR TYPICAL PRODUCTS.⁴³

A second anomaly is that, despite the positive correlation between patent effectiveness and the costs of imitating patented products, in several industries patents were relatively ineffective and duplication costs were nonetheless very high, whether or not the innovation was patented. Among these were guided missiles and several types of industrial machinery (food products machinery, electric welding apparatus, and speed changers, drives, and gears). In these instances the relative complexity of the products presumably makes reverse engineering inherently costly despite relatively weak patent protection.

It is interesting to compare our findings with those of Edwin Mansfield, Mark Schwartz, and Samuel Wagner, who studied the effects of patents on imitation costs in three industries,⁴¹ They concluded that

patents generally raised imitation costs by 30 percentage points in drugs, 20 points in chemicals, and 7 points in electronics. To render our data comparable, we evaluated each respondent's answer at the mean of the relevant range and computed crude industry average imitation costs for each type of innovation.⁴² Our results were consistent with those of Mansfield, Schwartz, and Wagner. We found that patents raise imitation costs by 40 percentage points for both major and typical new drugs, by 30 points for major new chemical products, and by 25 points for typical chemical products. In electronics, our results differed somewhat for semiconductors, computers, and communications equipment, but the range was 7 to 15 percentage points for major products and 7 to 10 for typical products.⁴³

Although the costs and time required for duplication are related to the effectiveness of patents, they do not seem to be linked strongly to any other mechanism of appropriability. In particular, most imitation time and cost measures are uncorrelated with lead-time and learning curve advantages, and where such correlations are statistically significant (at the level of the individual respondent), the correlation coefficient is invariably below .15. These results make sense. Lead-time and learning advantages may permit appropriation of returns even when duplication is relatively quick and inexpensive. Effective patents, however, presumably require considerable time and expense to be invented around.

Finally, most of our respondents believed only a few firms were capable of duplicating new processes and products. As table 10 shows, the median and modal number of firms judged capable of duplicating a major process or product innovation was three to five. The median and modal number of firms regarded as capable of duplicating a typical process or product innovation was six to ten. The data revealed only the slightest tendency toward a smaller number of capable duplicators for processes than for products.

R&D AND INNOVATION

In this section, we summarize how data derived from our survey have been employed to understand better the sources of interindustry differences in R&D spending and the rate of technological advance. In the first such effort Richard Levin, Wesley Cohen, and David Mowery used several survey-based measures to explain variations in the published Federal Trade Commission data on industry-level R&D spending as a percentage of sales.⁴⁴ They also sought to explain interindustry differences in the rate at which new processes and new products were introduced during the 1970s, as reported by our survey respondents.⁴⁵ In a subsequent paper, Cohen, Levin, and Mowery studied the extent to which the same survey-based measures explained the powerful industry effects in the confidential FTC data on R&D intensity at the level of the business unit.⁴⁶

The first paper focused on the Schumpeterian hypothesis that R&D intensity and the level of industry

FINALLY, MOST OF OUR RESPONDENTS BELIEVED ONLY A FEW FIRMS WERE CAPABLE OF DUPLICATING NEW PROCESSES AND PRODUCTS. AS TABLE 10 SHOWS, THE MEDIAN AND MODAL NUMBER OF FIRMS JUDGED CAPABLE OF DUPLICATING A MAJOR PROCESS OR PRODUCT INNOVATION WAS THREE TO FIVE. THE MEDIAN AND MODAL NUMBER OF FIRMS REGARDED AS CAPABLE OF DUPLICATING A TYPICAL PROCESS OR PRODUCT INNOVATION WAS SIX TO TEN. THE DATA REVEALED ONLY THE SLIGHTEST TENDENCY TOWARD A SMALLER NUMBER OF CAPABLE DUPLICATORS FOR PROCESSES THAN FOR PRODUCTS.

Table 10. Number of Firms Capable of Duplicating an Innovation, Frequency Distribution of Median Responses

<i>Type of innovation</i>	<i>None</i>	<i>1 or 2</i>	<i>3 to 5</i>	<i>6 to 10</i>	<i>More than 10</i>
Major new or improved process	2	32	75	18	2
Typical new or improved process	1	7	41	58	22
Major new or improved product	2	25	73	25	4
Typical new or improved product	1	5	33	63	26

Source: Survey of 129 lines of business.

concentration significantly influences innovation rates. One common rationale for this hypothesis is that industry concentration enhances the potential for appropriation of R&D returns. A different view is that, in the long run, concentration tends to be a consequence of industry evolution in a regime of abundant technological opportunity and a high degree of uncertainty associated with investment in R&D. Both perspectives suggest that there is no simple, causal relationship between concentrations per se and R&D. Concentration may be statistically significant in simple regression specifications because it reflects the influence of the unobserved appropriability and opportunity conditions that directly affect R&D spending and the rate of innovation.

In ordinary least squares and two-stage least squares specifications that included only the four-firm concentration ratio and its square as Regressors, Levin, Cohen, and Mowery replicated with the industry level FTC data the familiar inverted-U relationship between concentration and R&D intensity, and they found a strong relationship of the same form between concentration and the rate of innovation.⁴⁷ Adding two-digit industry fixed effects weakened slightly the effect of concentration on R&D, but the innovation-rate equation was unaffected.

The results changed dramatically with the addition of measures of appropriability and technological opportunity derived from the survey.⁴⁸ Whether or not two-digit industry fixed effects were included, the coefficients on concentration and its square fell by an order of magnitude in the R&D equation, and the effect of concentration was no longer statistically significant at the .05 levels in either the R&D intensity or the innovation-rate equation. The vector of survey-based opportunity variables was significant at the .05 level in all specifications, and the opportunity and appropriability variables were jointly significant. The appropriability variables, however, were not individually significant in the R&D equation, although the rate of innovation was positively related to the effectiveness of an industry's most effective means of appropriation.⁴⁹

The paper by Cohen, Levin, and Mowery used the disaggregated FTC data at the level of the business unit to investigate the Schumpeterian hypothesis linking size and R&D intensity. The authors found that when either fixed industry effects (at the level of the line of business) or survey based industry characteristics were taken into account, firm size had a very small and statistically insignificant effect on R&D intensity. The size of the business unit did have a significant effect on the probability of engaging in R&D, but there was no perceptible tendency for R&D intensity to increase with size within the group of R&D performers. Size effects, however, explained only two-tenths of 1 percent of the variance in R&D intensity, while industry effects at the line-of-business level explained half this variance.

Cohen, Levin, and Mowery found that industry-level measures of appropriability, opportunity, and demand conditions were consistently significant in ordinary least squares, GLS, and Tobit regressions explaining business unit R&D intensity. Moreover, these industry characteristics explained approximately half the variance in

R&D intensity explained by fixed industry effects. When attention was focused on those lines for which there were at least three survey respondents, measured industry characteristics explained 56 percent of the variation in R&D intensity among industries. Within particular two-digit industries (chemicals, machinery, and electrical equipment), measured characteristics explained 78 to 86 percent of the variance explained by fixed effects.

The results obtained in the two papers indicated that survey-based measures could contribute substantially to an explanation of interindustry differences in R&D intensity and innovative performance. Measures derived from the survey, despite their imperfections, have also been found useful for various other purposes.⁵⁰

REMARKS ON POLICY

Our findings suggested some general principles relevant to policies that affect the incentives to engage in innovative activity.

A first principle is that the patent system and related institutions to protect intellectual property should be understood as social structures that *improve* the appropriability of returns from innovation. They are not the only nor necessarily the primary barriers that prevent general access to what would otherwise be pure public goods. Lead-time accrues naturally to the innovator, even in the absence of any deliberate effort to enhance its protective effect. Secrecy, learning advantages, and sales and service efforts can provide additional protection, though they require the innovator's deliberate effort. The survey confirmed that these other means of appropriation are typically more important than the patent system. Hence in examining a proposed adjustment of the patent system or related institutions, it is important to recognize that the incremental effect of the policy change depends on the protection other mechanisms provide.

A FIRST PRINCIPLE IS THAT THE PATENT SYSTEM AND RELATED INSTITUTIONS TO PROTECT INTELLECTUAL PROPERTY SHOULD BE UNDERSTOOD AS SOCIAL STRUCTURES THAT *IMPROVE* THE APPROPRIABILITY OF RETURNS FROM INNOVATION.... THE SURVEY CONFIRMED THAT THESE OTHER MEANS OF APPROPRIATION ARE TYPICALLY MORE IMPORTANT THAN THE PATENT SYSTEM....FROM THIS FOLLOWS OUR SECOND MAJOR PRINCIPLE, WHICH IS THAT THE INCREMENTAL EFFECTS OF POLICY CHANGES SHOULD BE ASSESSED AT THE INDUSTRY LEVEL.

The survey results also confirmed substantial interindustry variation in the level of appropriability and in the mechanisms that provide it. From this follows our second major principle, which is that the incremental effects of policy changes should be assessed at the industry level. For example, in the aircraft industry, where other mechanisms provide considerable appropriability, lengthening the life of patents would tend to have little effect on incentives for innovation. In the drug industry the effect of a longer lifetime would matter more.⁵¹

Finally, improving the protection of intellectual property is not necessarily socially beneficial. Empirical work has so far indicated a positive cross-sectional relationship between strong appropriability, as measured by variables constructed from our survey, and innovative performance. But the social cost-benefit calculation is not straightforward. Stronger appropriability will not yield more innovation in all contexts and, where it does, innovation may come at excessive cost.

To illustrate how our survey results and general perspective might inform policy discussion, consider the

1987 proposal (S. 438, H.R. 557) that patent license agreements and other contracts relating to the use of intellectual property “shall not be deemed illegal per se under any of the antitrust laws.” One consequence would be to eliminate the per se illegality of tie-in arrangements (those in which purchase of one product, the “tying product,” is dependent on purchase of other products) where the tying product is covered by a patent or otherwise protected as intellectual property.⁵² Our findings have suggested some issues a court should consider in evaluating such a tying arrangement under the rule of reason.

When the rule of reason is applied to tying cases, a relevant consideration is the firm’s power in the market for the tying good. Courts have often presumed that intellectual property protection is itself evidence for such

FINALLY, IMPROVING THE PROTECTION OF INTELLECTUAL PROPERTY IS NOT NECESSARILY SOCIALLY BENEFICIAL. EMPIRICAL WORK HAS SO FAR INDICATED A POSITIVE CROSS-SECTIONAL RELATIONSHIP BETWEEN STRONG APPROPRIABILITY, AS MEASURED BY VARIABLES CONSTRUCTED FROM OUR SURVEY, AND INNOVATIVE PERFORMANCE. BUT THE SOCIAL COST-BENEFIT CALCULATION IS NOT STRAIGHTFORWARD. STRONGER APPROPRIABILITY WILL NOT YIELD MORE INNOVATION IN ALL CONTEXTS AND, WHERE IT DOES, INNOVATION MAY COME AT EXCESSIVE COST.

power. To the other good reasons for rejecting such a presumption,⁵³ we add that the mere existence of a patent or other legal protection says nothing about its efficacy in a competitive context. As the survey results showed, the effectiveness of protection varies widely among industries. Thus in deciding a case, a court should inquire into the actual competitive significance of intellectual property protection in the particular market.

Suppose, for example, that a pharmaceutical company were to tie hospital sales of supplies or equipment to its sale of a patented drug. Since patent protection of drugs is generally strong and effective, and a drug is often uniquely suited for particular purposes, skepticism about the reasonableness of the tie-in would be in order. The arrangement could not plausibly be regarded as a straightforward means of appropriating returns to which the firm was entitled as owner of the patent. Given the typical effectiveness of drug patents,

the price of the drug should suffice for that purpose. There might, of course, be benign explanations for the tie; for example, if the supplies or equipment were complementary to the use of the drug, the arrangement might be explicable as an attempt to control the quality of treatment. But if the evidence supported no such explanation, the tie would seem an unreasonable restraint of trade.

By contrast, consider a producer of a patented product in an industry where no mechanism of appropriability functions particularly well plywood, for example, where patents, secrecy, lead time, and learning advantages are all rated no higher than four on a seven-point scale of effectiveness. In this instance the low level of appropriability in general and the ineffectiveness of patents in particular should weigh against any presumption that a patent confers market power. The patentee in such an industry should be entitled to some scope for ingenuity in constructing arrangements that maximize the return to the patent, provided that these arrangements are not open to antitrust objections on grounds independent of the role played by the patent.

The intellectual property provisions of the Omnibus Trade and Competitiveness Reform Act also serve to illustrate the relevance of the survey results. One provision requires the US trade representative to identify

countries that have been particularly insensitive, as a matter of law or de facto policy, to the need for protection of intellectual property and to initiate unfair trade practice (section 301) investigations against them.⁵⁴ This provision of the trade bill would complement the administration's diplomatic throughout the world and particularly in countries that permit firms to copy patented or copyrighted products from the United States.

Since the impact of legal protection of intellectual property depends on the strength of other appropriability mechanisms and varies widely among industries, focused efforts to solve problems in specific markets would be more prudent than a broad attempt to upgrade protection. There is little point in expending diplomatic capital to compel foreign countries to pass or enforce laws that, in most industries, would have minimal impact on the competitive process. By contrast, in those specific industries such as pharmaceuticals-in which patent protection is effective, other means of appropriation are poor substitutes, and foreign governments often restrict, officially or tacitly, the ability of U.S. firms to exploit patents-a more persuasive case could be made for the United States to pressure its trading partners to change their behavior.

TO THE OTHER GOOD REASONS FOR REJECTING SUCH A PRESUMPTION,⁵³ WE ADD THAT THE MERE EXISTENCE OF A PATENT OR OTHER LEGAL PROTECTION SAYS NOTHING ABOUT ITS EFFICACY IN A COMPETITIVE CONTEXT. AS THE SURVEY RESULTS SHOWED, THE EFFECTIVENESS OF PROTECTION VARIES WIDELY AMONG INDUSTRIES. THUS IN DECIDING A CASE, A COURT SHOULD INQUIRE INTO THE ACTUAL COMPETITIVE SIGNIFICANCE OF INTELLECTUAL PROPERTY PROTECTION IN THE PARTICULAR MARKET.

APPENDIX: DETAILS OF SAMPLE CONSTRUCTION

Our review of the FTC data indicated that several lines of business did not report any R&D activity, and several others were aggregated to prevent violating confidentiality rules. Anticipating difficulty in finding knowledgeable respondents in industries without formal R&D activity, and wishing to avoid industry categories that included technologically disparate products, we eliminated those lines of business from our sampling frame.

The industries eliminated on grounds of heterogeneity were either the FTC's aggregations of technologically disparate industries or those corresponding to SIC industries with four-digit codes ending with 9. Such industries are residual categories within the relevant three-digit groups; their titles usually contain the words "miscellaneous, not elsewhere classified."

Confidentiality requirements prohibited us from using the FTC data as a means of identifying the firms that conduct R&D in each line of business. Instead, we used the *Business Week* annual R&D survey to identify all publicly traded firms that reported R&D expenses in excess of either 1 percent of sales or \$35 million. This constitutes a nearly comprehensive list of private firms performing significant R&D. There were 746 such firms in 1981, when our survey design efforts commenced.

We used the information in Dun and Bradstreet's *Million Dollar Directory* to assign each of the *Business Week* firms to its major lines of business. Dun and Bradstreet's does not provide a complete list of each firm's lines of business, but it indicates as many as six four-digit SICs for each firm, in rough order of sales. Since

SINCE THE IMPACT OF LEGAL PROTECTION OF INTELLECTUAL PROPERTY DEPENDS ON THE STRENGTH OF OTHER APPROPRIABILITY MECHANISMS AND VARIES WIDELY AMONG INDUSTRIES, FOCUSED EFFORTS TO SOLVE PROBLEMS IN SPECIFIC MARKETS WOULD BE MORE PRUDENT THAN A BROAD ATTEMPT TO UPGRADE PROTECTION. THERE IS LITTLE POINT IN EXPENDING DIPLOMATIC CAPITAL TO COMPEL FOREIGN COUNTRIES TO PASS OR ENFORCE LAWS THAT, IN MOST INDUSTRIES, WOULD HAVE MINIMAL IMPACT ON THE COMPETITIVE PROCESS.

some firms operate in nonmanufacturing industries, in manufacturing industries absent from our sample, or in two or more industries that fall into only one FTC line of business, we had substantially less than 746 x 6 observations. Within our sample lines of business, we found a total of 1,928 units operated by 688 firms.

A major design problem was how to obtain responses for business units within the same firm. Of our 688 firms, 470 participated in more than one of our sample lines of business. We initially attempted to identify relevant respondents using *Industrial Research Laboratories of the United States*. But our pretest subjects told us that more than half the people in such a sample were inappropriate. Some had been assigned to the wrong line of business; others had been promoted or had left the relevant division or the firm.

We therefore adopted a two-stage approach in which each firm's senior R&D vice president or chief executive officer was asked to furnish the names of

employees with the knowledge to complete the questionnaire for specific lines of business. We sent first-round requests to 470 firms representing 1,710 business units. There was attrition of 332 business units from this sample for three reasons: the firm did not do R&D in the specified line of business, the industry definition did not fit any of its activities, or a respondent could not be located. From this adjusted sample frame of 1,378 business units in firms with multiple units, we received names of respondent's for 716. We sent questionnaires to each of these potential respondents as well as to representatives of the 218 firms operating in only one line of business. At this stage, there was some further attrition in the sample. Ultimately, we received 650 completed questionnaires from an overall adjusted sample frame of 1,562-an overall response rate of 41.6 percent.

COMMENT AND DISCUSSION

Richard Gilbert: The authors' research program will have lasting value for people interested in R&D markets and markets for intellectual property. They are correct in focusing on appropriability as a key factor in the incentive to undertake R&D. And their findings are generally consistent with those of other studies, for example, those by Christopher T. Taylor and Z. A. Silberston and Edwin Mansfield, Mark Schwartz, and Samuel Wagner.¹ While this consistency may take a bit of the drama out of what Levin and his colleagues have done, the convergence of knowledge on this subject gives us some reason to believe we might be getting to the truth.

One of the authors' main conclusions is that there are very large differences, both among industries and within them, in the effectiveness of various means of appropriating intellectual property and also in the cost of imitation. It is an important result, but one that may cause some consternation. A main function of micro-economic theory is to form testable generalizations about the way the world works. Some of the work implies - at least, in the market for R&D and intellectual property - that such generalizations are extremely risky. We

might be inventing a new field of microeconomic analysis, or “picoeconomics.” Picoeconomics would keep us busy for a long, long time. But if we go down that path, our models will soon become as complicated as the world we are trying to explain.

These particular authors do not seem to want to lead us toward picoeconomics. It is apparent from their follow-up work that they intend to draw some general conclusions about how appropriability varies across industries. Clearly, we would like to know how market structure and capital intensity in different industries influence the degree of appropriation and affect incentives to innovate. But merely adding appropriation as another explanatory variable in these regressions does little good. We really want to know whether there are systematic relations between the degree of appropriation and other observable economic variables.

With regard to methodology, I suggest using a weighting scheme based on the amount of R&D a firm does and, perhaps, the number of patents it has produced. The purpose would be to weight responses according to the quality of the information. Some industries seem not to have performed any R&D for twenty years. Although it is important to know why these firms have not been active, their responses should be adjusted to reflect the information they possess. Also, I suggest that in their survey work the authors include a definition of R&D. That is not a trivial task because there is much variation in what is called research and development.

I would have liked the survey to address more directly some of the theoretical issues in the economics of R&D. Various models in the R&D literature have different implications for the simultaneous determination of R&D intensity and market structure. For example, models such as that of Partha Dasgupta and Joseph Stiglitz imply that the current rate of R&D spending should be independent of cumulative R&D expenditures by a firm or the rivals of the firm.² This is a consequence of the constant-hazard rate model. Other researchers such as Drew Fudenberg and his colleagues imply that past R&D is crucial to current and future R&D expenditures.³ The dynamic implications of these models are very different. The preemption-type models also suggest that a history of successful R&D gives a firm a technological advantage that provides some protection from future R&D competition and tends to increase concentration in a market.

I would be interested in knowing if the survey could have elicited some kind of response about the way R&D success alters the competitive environment of the firms and, conversely, how the competitive environment influences R&D spending. The questions in which the investigators ask how many firms are viable competitors with a given firm and how many could replicate its R&D bear on this question. It is interesting that the number of serious rivals for each firm was small, somewhere between three and six.

There is an empirical problem with surveys of the relationship between competition and R&D. If R&D really does have an effect on entry and competition, then the sample is necessarily biased. There are potential competitors who were not represented in the sample because the firms failed. How one accounts for the failures and puts them back in the sample is a difficult empirical problem.

In terms of patent policy, the diversity and effectiveness of patent protection across industries raises an obvious and interesting question. How should patent policy deal with the large differences in the values of patents among various industries? The authors point to the example of the semiconductor industry and suggest that the industry need not be any worse off as a consequence of the limited patent protection it has received. But one might also question whether the performance of the pharmaceuticals industry might have been improved if patent protection had been circumscribed to some extent.

I have experimented with a very simple model of optimal patent life with limited appropriability. It is basically a Nordhatis-type model with entry and spillovers.⁴ It shows that the optimal patent life is not a function of the size of the innovation, so one does not have to worry that there are big innovations in some industries and small innovations in others. The optimal patent life does, however, depend on elasticities of R&D and its benefits and costs. It also depends on the degree of appropriability, and there is the intuitive answer that the optimal life is inversely related to the extent of private appropriation of the social value of the invention. This result suggests that we need either more protection in semiconductors or less protection in the patent drug industry.

The authors' survey provides a basis for contrasting patents with other approaches to protect intellectual property. A patent is a peculiar policy instrument. It represents an unnatural barrier to market entry that is erected to facilitate private appropriation. The survey suggests that other factors may be more important as a means of appropriation, and that other unnatural barriers might be more effective in stimulating R&D. As an extreme example (which I am not proposing), a tax on capital could arguably make entry more difficult and therefore stimulate R&D. It would be useful to do a survey of the effectiveness of different laws governing rights for intellectual property in different countries in an industry such as pharmaceuticals, which is one of the few in which patents really do seem to matter.

The survey results suggest that patents are important as a barrier to entry in the semiconductor industry not because they protect an individual innovation or invention but because they provide a hurdle for potential entrants, who have to acquire a package of marketable processes and products that they can cross-license to other firms. While this seems an inappropriate or at least unintended outcome of the patent grant, it might well be that by increasing ordinary barriers to entry in the semiconductor industry, the returns to research and development would be enhanced. This is another illustration of the Schumpeterian hypothesis and the tensions between strong enforcement of the antitrust laws and the desire to provide a stable platform for encouraging investment in R&D. If these observations with regard to the semiconductor industry generalize to other markets, they provide a starting point for further reexamination of the antitrust laws in the context of industrial R&D policy.

Thus this survey has raised some very interesting questions. Now we have to get on with their resolution. Thanks to this project, we have some of the data we need for the job.

Zvi Griliches: We should be grateful to Richard Levin and his associates for providing us with a new and detailed glimpse into a subject that is both very important and also lacking in good data. Far too little fresh economics data is collected, and we all have much to learn from the effort of this endeavor. That I am going to quarrel with some of the authors' assessments does not diminish in my eyes the basic value of this enterprise.

The authors have collected a large set of responses from many individuals located in different industries. This multipurpose survey will have many uses as we learn more about the responses and how to interpret them. I will focus on how these answers can help us learn which industries find patents an effective mechanism for appropriating returns from innovative effort, which ones do not, and whether mechanisms are available instead of or in addition to patents.

"Conditions of appropriability" determine the returns from a given innovative effort and hence the incentive to engage in it. One would expect that in industries in which appropriability is easy, there would be more innovative effort, higher returns, and a faster rate of technological progress. Such conditions may not be a fixed,

unchanging characteristic of an industry, however. As more inventive effort is pursued, projects may become less easily appropriable, information may be leaked, and conditions may actually equalize among industries. The problem here is the same as in most empirical research programs: What is exogenous and what is endogenous?

There are two generic problems with using the responses from such a survey: Are the responses comparable among individuals and do they reflect real differences among industries? Given the use of a scale of one to seven, I remain unsure about whether one person's response of five is equivalent to another's of four or six. Most questions do not have an objective anchor and could, therefore, differ greatly in the meanings attached to them by different respondents. This may account for the large dispersion in responses to most questions even within the same industry. It also leads to the difficulty of deciding whether the responses reflect real differences across

Table 1. Analysis of Variance of Differences among Industries in Responses to Questions on the Effectiveness of Different Appropriability Mechanisms

Question	Individual response ^a		Aggregated industry response ^b	
	Processes	Products	Processes	Products
Patents to prevent duplication	1.9	2.8	3.4	5.5
Patents to secure royalties	1.6	1.4	2.4	2.3
Secrecy	1.2	1.2	1.3	1.3
Lead time	1.7	1.6	2.2	2.5
Quickly down the learning curve	1.0	1.0	.9	1.2
Superior sales or service	1.2	.9	.7	.9
Average of questions 1 and 2		2.1		4.8
Average of questions 3 through 6		1.1		1.6

Source: Authors' calculations.
a. Covers 541 responses for 130 industries; approximate .05 significance level is 1.3.
b. Covers 620 responses for 24 industries; approximate .05 significance level is 1.6.

industries or just random fluctuations among individuals.

There is a surprising amount of variability within industries in responses to the same question. Some questions, such as those on the effectiveness of patents, are reasonably objective and seem to have a variance among industries. Other questions, such as whether secrecy is effective, do not seem to be particularly industry-specific and do not discriminate well among industries. Questions about lead-time, secrecy, sales effort, and service quality are really questions about different ways of succeeding, not about properties of an industry. It is well to have a long lead-time or to achieve secrecy, but how is that to be accomplished? These are not characteristics over which either the firm or the policy makers have clear control. Patents are at least a somewhat better defined instrument, and we have some ideas about how the patent system could be tinkered with.

Table 1 presents my analysis of variance results for some of the responses derived from the Yale survey. (I am grateful to Levin for providing me with the original survey data.) It shows that there is more variance among industries in the responses to questions on the effectiveness of patents, especially for products, and very little variance in the other questions on conditions for appropriability, especially for process innovations. Process innovations are clearly less industry-specific

IT SHOWS THAT THERE IS MORE VARIANCE AMONG INDUSTRIES IN THE RESPONSES TO QUESTIONS ON THE EFFECTIVENESS OF PATENTS, ESPECIALLY FOR PRODUCTS, AND VERY LITTLE VARIANCE IN THE OTHER QUESTIONS ON CONDITIONS FOR APPROPRIABILITY, ESPECIALLY FOR PROCESS INNOVATIONS. PROCESS INNOVATIONS ARE CLEARLY LESS INDUSTRY-SPECIFIC AND SO IS THE IMPORTANCE OF SUPERIOR SALES AND SERVICE EFFORTS.

and so is the importance of superior sales and service efforts.

Another way of seeing this problem is to look at the authors' table 6, which shows that for many of the nonpatent mechanisms the cross correlation among methods of appropriation is lower at the aggregated industry level than at the level of the individual response. If industrial classification mattered, one would expect higher correlations for the aggregated variables.¹ A very simple model illustrates this point. Assume that two questions effectively measure the same thing. Then a variance components model for responses to these questions would be

$$y_{qij} = m_i + a_j + e_{qij}$$

where m_i is the "true" industry effect perceived by all individuals, a_j is the individual deviation from the average respondent independent of the industry he is in, and e_{qij} is the random response error associated with the particular question q and individual ij . Taking these effects as independent from each other, and assuming that the first component does not average out as one aggregates within industries but that the other ones do in proportion to the average number of respondents per industry, gives us a little model that can be fit to the observed variances and covariances at the micro and macro industry levels. The following material shows the relevant numbers for two pairs of questions: whether product patents are effective against duplication and in securing royalty income, and whether moving quickly down the learning curve and superior sales and service efforts are effective in protecting the competitive advantage of new products. The two-question expected variance covariance matrix is

$$\begin{array}{cc} \textit{Individual level} & \textit{Industry level} \\ \left[\begin{array}{cc} \sigma_m^2 + \sigma_a^2 + \sigma_1^2 & \sigma_m^2 + \sigma_a^2 \\ \dots & \sigma_m^2 + \sigma_a^2 + \sigma_2^2 \end{array} \right] & \left[\begin{array}{cc} \sigma_m^2 + \frac{(\sigma_A^2 + \sigma_1^2)}{N_i} & \sigma_m^2 + \sigma_A^2/N_i \\ \dots & \sigma_m^2 + \frac{(\sigma_A^2 + \sigma_2^2)}{N_i} \end{array} \right] \end{array}$$

where m_i is the "true" industry effect perceived by all individuals, a_j is the individual deviation from the average respondent independent of the

	<i>Product patents</i> <i>IB1</i>	<i>Effective</i> <i>IB2</i>	<i>Learning curve</i> <i>IB5</i>	<i>Secrecy</i> <i>IB6</i>
Individual (<i>N</i> = 643)	2.860 ...	1.435 2.748	1.810803 1.662
Industry (<i>N</i> = 24)	.618384 .376	.148013 .075
Implied estimates (<i>N</i> = 27)				
σ_A^2		1.09		.820
σ_m^2		.34		<0

The numbers imply that the common variance between industries accounts for about one-eighth of the variance at the level of the individual response and more than half at the aggregated industry level. For the two other questions the implied "true" variance between industries is negative. (For the patents question the correlation rises from .51 at the level of the individual respondent to .80 for averages at the NSF industry level, while for the two appropriability questions the numbers go from .46 to .12, implying that such averaging at-

Table 2. Stock Market's Valuation of R&D and Patents, 722 U.S. Manufacturing Corporations, 1980^a

Variables	Coefficients						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>SP/A</i> ^b	.165 (.100)	.380 (.171)	.107 (.167)	.249 (.155)	.360 (.170)	.077 (.183)	.199 (.161)
<i>K/A</i> ^c932 (.201)	.335 (.178)898 (.224)	.342 (.175)
<i>NR</i> ^d	11.96 (1.37)	12.24 (1.38)
<i>PPP</i> ^e034 (.024)	.019 (.024)	.019 (.023)	.035 (.024)	.023 (.025)	.019 (.024)
<i>PPP · (SP/A)</i>236 (.116)	.075 (.110)	.098 (.101)	.267 (.133)	.115 (.142)	.164 (.128)
<i>PPP · (K/A)</i>365 (.130)432 (.172)	...
<i>PPP · NR</i>	2.788 (1.231)	2.60 (1.39)
<i>NPP</i> ^f039 (.079)	.100 (.089)	.054 (.075)
<i>NPP · (SP/A)</i>127 (.388)	.174 (.432)	.290 (.293)
<i>NPP · (K/A)</i>263 (.636)	...
<i>NPP · NR</i>	-1.89 (4.77)
\bar{R}^2	.166	.172	.200	.310	.170	.198	.309

Source: Adapted from Iain Cockburn and Zvi Griliches, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents," working paper 2465 (National Bureau of Economic Research, December 1987), tables 3b and 5.

a. Dependent variable is $\log Q$ (market value divided by replacement value). All equations contain also ten 2-digit SIC industry dummy variables and a logarithm of total assets variable whose coefficient is small but consistently significant, on the order of $-.03$ (.01).

b. Stock of patents (30 percent depreciation rate) divided by total net assets.

c. Cumulated R&D "capital" stock (15 percent depreciation rate) divided by total net assets.

d. Net R&D investment divided by net assets ($R - .15K/A$).

e. Sum of responses to "patents provide protection against duplication" questions for both process and product innovations. Averages at a fifty-five industries (approximately 3.5 digit level) aggregation.

f. Average of responses to all other "effectiveness of nonpatent appropriability mechanisms" questions.

tenuates rather than strengthens the relationship between the responses to such questions.) In short, while these questions might be interesting, they do not seem to be able to pick out significant differences among industries. Coming quickly down the learning curve and providing superior services are about equally effective.

Do the results of this survey help us explain other phenomena besides the relationship between answers to different sets of related questions within the survey? Looking at the responses to the questions evaluating patents and other appropriability mechanisms, the evidence appears to be mixed. Levin, Cohen, and Mowery did not find the appropriability variables significant in explaining differences among industries in R&D intensity, even in the absence of industry dummies.² In a forthcoming paper Iain Cockburn and Zvi Griliches use the Yale survey responses, aggregated to fifty-five industries at approximately a 3-digit SIC level, to see whether the stock market values the accumulated patents and the current R&D policy of a firm more or less in industries where the appropriability conditions are better in some sense.³ Table 2 reproduces typical results from this study. Patent effectiveness measures help in some sense. The equations seem to imply that both accumulated past patents and current R&D moves are valued more by the market when patent protection is effective. Other appropriability measures do not help. But neither set of measures does better than just an interaction with ten higher-level (2-digit) industries dummies. The greater detail available in the Yale survey appears to be counterbalanced by the (inevitably?) greater imprecision of these measures at the detailed industry level. So there is something there but not as much as might be wished. But we should be thankful for there is hope that a more detailed study of these and other responses in this survey will help us understand our world better. In particular, the information on the differential connectedness of science in different industries is very intriguing and may be of help in

future analyses of the contribution of science to technological advance.

GENERAL DISCUSSION

Richard Levin agreed with Zvi Griliches that the appropriability variables could not discriminate effectively among more than about ten industry groupings, but he suggested that this may be a good thing, especially in light of Richard Gilbert's concern that studies such as this amount to "picoeconomics," from which no generalizations can be drawn. Sidney Winter noted that the results suggest there may be a relatively short list of variables to consider in an analysis of appropriability and incentives for R&D, and that ten industry groupings may provide all the information needed. In other dimensions, a finer division may be important. For example, Levin pointed out that, relative to questions on lead time, learning curves, and duplication costs, the questions on patent effectiveness discriminate better, as do the questions on learning and information spillover and those on duplication time.

Moreover, headed, a principal conclusion, that patents do not matter very much except in the chemical industries and in semiconductors, comes through regardless of problems with questions about other mechanisms of appropriation. In these two industries, the meaning and role of patent protection is different. Chemical products are easy to patent because the structure of the molecule of each product is unique, but patents are easy to invent around because it is often possible to create a discrete but structurally similar product with similar properties. With semiconductors, however, the innovation process is cumulative, with each invention built very distinctly on the previous one. The innovation provided by one firm makes the product invented by another firm more valuable. So the role patents play is to define the property rights (usually through the licensing process) so that the proceeds of this cumulative process can be shared and innovation can be encouraged.

Griliches also raised questions about whether the variables measured in this study are appropriately regarded as exogenous. Sales and service effort, for example, is a choice variable for the firms, and hence effectiveness should be endogenous. Likewise, lead time should be thought of as an outcome of the technology race, rather than an exogenous condition of it. Winter pointed out that even the distinction of product and process may be somewhat endogenous in the sense that firms take into consideration the importance of secrecy and the possibility of reverse engineering in designing their products. Firms often work to make their high-technology products inaccessible to reverse engineering, he noted, which tends to make these products more like processes from the standpoint of appropriability.

Several participants seemed concerned about problems of measurement and scaling biases in the data. Levin responded that various techniques, such as weighting the responses by the inverse of the variances or standard deviations of the individual responses, were tried to correct for these biases. The principal findings were robust to efforts to stretch or squeeze the distribution of responses, he noted, but it was unclear what such "corrections" mean since no one knows what the true distribution should be.

Joseph Farrell took issue with Gilbert's argument that it is not particularly interesting or helpful to worry about the determinants of R&D in industries that do not perform R&D. In fact, he suggested, it would be very useful to know why some industries seem to do so little research while others do so much. Gilbert agreed, but argued that it was still important to assign some sort of weights to individual responses

to particular questions, based on the respondent's experience with those issues. Martin Baily argued that weighting the responses by how much R&D the responding firms do would be inappropriate, however, since the amount of R&D is what the authors are trying to explain.

Richard Schmalensee suggested that some of the within-industry variance in the responses about sales and service and other mechanisms of appropriation may be due to the fact that the R&D executives who responded to the questionnaire are less knowledgeable about what happens to the product after it leaves their jurisdiction in the organization.

Responding to a question from Paul Joskow, Levinsaid if he were doing the survey over again, he would want to do more pretesting. For example, the authors might have picked up on the issue of intraindustry variance earlier if they had pretested multiple respondents from the same industry. Or they might have learned ways to restructure questions to discriminate more carefully between exogenous and endogenous factors.

Robert Litan noted that one implication of the study for trade policy is that issues of intellectual property rights should be dealt with industry by industry. This is, in effect, how section 301 of the Trade Act already works, he added. The section provides a procedure for pursuing complaints about unfair trade practices abroad, but these complaints must be brought product by product. Litan also suggested that mechanisms of appropriating returns from R&D might vary between large and small firms. Levin agreed, but noted that the sampling procedure surveyed only publicly held firms, so that start-up ventures were completely excluded. Patents may be much more important for a start-up company because they provide something tangible to sell if the firm tries to sell out later. ▲

We are grateful for the support of the National Science Foundation and especially to Rolf Piekartz of the NSF's Division of Policy Research and Analysis.

We also wish to thank the 650 respondents to our survey and the R&D executives who helped us pretest it—especially Ralph Gomory, Bruce Hannay, and Lowell Steele. Donald DeLuca, Wendy Horowitz, and other members of the Roper Center for Survey Research helped manage the survey. Robert W. Wilson, and Margaret Blair, Marc Chupka, Emily Lawrance, Constance Helfat, Andrew Joskow, Kathleen Rodenrys, Somi Seong, Andrea Shepard, and Hal VanGieson also provided valuable assistance.

1. F. M. Scherer and others, *Patents and the Corporation: A Report on Industrial Technology under Changing Public Policy*, 2d ed. (privately published, 1959); and C. T. Taylor and Z. A. Silberston, *The Economic Impact of the Patent System: A Study of the British Experience* (Cambridge University Press, 1973).

2. See, for example, Meir Statman, "The Effect of Patent Expiration on the Market Position of Drugs," in Robert B. Helms, ed., *Drugs and Health: Economic Issues and Policy Objectives* (Washington, D.C.: American Enterprise Institute, 1981), pp. 140-51.

3. The importance of lead time and learning curve advantages is documented in Almarin Phillips, *Technology and Market Structure: A Study of the Aircraft Industry* (Lexington Books, 1971); and John E. Tilton, *International Diffusion of Technology: The Case of Semiconductors* (Brookings, 1971). For the importance of marketing and customer service, see Marie-Therese Flaherty, "Field Research on the Link between Technological Innovation and Growth: Evidence from the International Semiconductor Industry," working paper 84-83 (Harvard University, Graduate School of Business Administration, no date).

4. See, especially, Ariel Pakes, "Patents as Options: Some Estimates of the Value of

5. Holding European Patent Stocks," *Econometrica*, vol. 54 (July 1986), pp. 755-84. For a summary of the best of this work, see Zvi Griliches, Ariel Pakes, and Bronwyn H. Hall, "The Value of Patents as Indicators of Inventive Activity," working paper 2083 (Cambridge, Mass.: National Bureau of Economic Research, November 1986). For other perspectives on the usefulness of patent data, see the special issue of *Research Policy*, vol. 16 (August 1987).

6. F. M. Scherer, "The Propensity to Patent," *International Journal of Industrial Organization*, vol. I (March 1983), pp. 107-28; and John Bound and others, "Who Does R&D and Who Patents?" in Zvi Griliches, ed., *R&D, Patents and Productivity* (University of Chicago Press for National Bureau of Economic Research, 1984), pp. 21-54.
7. Richard C. Levin, "Toward an Empirical Model of Schumpeterian Competition," working paper 43 (Yale University, School of Organization and Management, 1981); Richard C. Levin and Peter C. Reiss, "Tests of a Schumpeterian Model of R&D and Market Structure," in Griliches, ed., *R&D, Patents and Productivity*, pp. 175-204; and Ariel Pakes and Mark Schankerman, "An Exploration into the Determinants of Research Intensity," in Griliches, ed., *R&D, Patents and Productivity*, pp. 209-32.
8. See Computer Software Act of 1980; Semiconductor Chip Protection Act of 1984; *Diamond v. Chakrabarty*, 447 U.S. 305 (1980), holding that plant and animal life is patentable under U.S. patent law; and D. J. Quigg, memorandum of April 7, 1987, explaining the policies of the U.S. Patent and Trademark Office concerning applications to patent life forms.
9. See H.R. 3, the Omnibus Trade and Competitiveness Reform Act of 1987, which is currently under consideration by a House-Senate conference committee.
10. H.R. 557 and S. 438, 100 Cong., 1 sess.
11. This is the "free access" externality, first emphasized in the context of innovation in Yoram Barzel, "Optimal Timing of Innovations," *Review of Economics and Statistics*, vol. 50 (1968), pp. 348-55. For a survey of the literature on patent races, see Jennifer Reinganum, "The Timing of Innovation: Research, Development and Diffusion," in Richard Schmalensee and Robert Willig, ed., *Handbook of Industrial Organization* (North Holland, 1988).
12. Richard R. Nelson, "Assessing Private Enterprise: An Exegesis of Tangled Doctrine," *Bell Journal of Economics*, vol. 12 (Spring 1981), pp. 93-111; and William D. Nordhaus, *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (MIT Press, 1969).
13. Richard C. Levin, "The Semiconductor Industry," in Richard R. Nelson, ed., *Government and Technical Progress: A Cross-Industry Analysis* (Pergamon Press, 1982), pp. 9-100.
14. Among the sources of ideas for the questions are Paul Allan David, *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century* (Cambridge University Press, 1975); Richard R. Nelson and Sidney G. Winter, "In Search of Useful Theory of Innovation," *Research Policy*, vol. 6 (Winter 1977), pp. 36-76; Nathan Rosenberg, "Science, Invention and Economic Growth," *Economic Journal*, vol. 84 (March 1974), pp. 90-108; and Devandra Sahal, *Patterns of Technological Innovation* (Addison-Wesley, 1981). For empirical literature on the economic effects of the patent system, see Scherer and others, *Patents and the Corporation*; and Taylor and Silberston, *The Economic Impact of the Patent System*. For imitation costs, see Edwin Mansfield, Mark Schwartz, and Samuel Wagner, "Imitation Costs and Patents: An Empirical Study," *Economic Journal*, vol. 91 (December 1981), pp. 907-18.
15. These managers had experience in communications equipment, industrial inorganic chemicals, metal cutting machine tools, shoe machinery, household electrical appliances, processed foods, computing equipment, semiconductors, copper smelting and refining, radio and TV sets, and industrial organic chemicals. They were asked to complete the questionnaire for a specific line of business, but to keep in mind the suitability of the questions for other lines of business with which they were familiar. After completing the questionnaire, they were interviewed face-to-face or by telephone. Interviews typically lasted one-half hour or more, and each question was discussed to eliminate sources of ambiguity.
16. The questions were similar to those in Mansfield, Schwartz, and Wagner, "Imitation Costs and Patents," but covered typical rather than specific innovations. Our industry sample was also broader.
17. One objective was to examine "natural trajectories" of the sort described in Nelson and Winter, "In Search of Useful Theory," p. 56.
18. Data from responses to questions in parts 3 and 4 have been used in Richard C. Levin, Wesley M. Cohen, and David C. Mowery, "R&D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review*, vol. 75 (May 1985, *Papers and Proceedings*, 1984), pp. 20-24; Cohen, Levin, and Mowery, "Firm Size and R&D Intensity: A Re-examination," *Journal of Industrial Economics*, vol. 35 (June 1987), pp. 543-65; and Richard C. Levin and Peter C. Reiss, "Cost-Reducing and Demand-Creating R&D with Spillovers" (Stanford University, Graduate School of Business, 1986). See Richard R. Nelson, "Institutions Supporting Technical Advance in Industry," *American Economic Review*, vol. 76 (May 1986, *Papers and Proceedings*, 1985), pp. 186-89, for findings on the importance of external sources of technological knowledge.

19. F. M. Scherer, "Inter-industry Technology Flows in the United States," *Research Policy*, vol. II (August 1982), pp. 227-45.
20. Interindustry differences are significant at the 0.05 level for approximately 60 percent of the questions in parts I and 2 of the questionnaire. If a higher level of aggregation is used to measure industry effects, such as the level at which the National Science Foundation reports R&D spending (a hybrid of two- and three-digit level industries), interindustry differences are significant at the 0.05 level for 70 percent of the questions.
21. Heterogeneity, as anthropologists have long insisted, is, however, in the eye of the beholder. One R&D manager, asked to inform us about the air and gas compressor industry, inquired whether we were interested in large, medium, or small compressors. In his view the technologies were fundamentally different. We asked him to note on his questionnaire where the answers to our questions differed across these size categories. The booklet he returned contained no such notation.
22. One notable consequence arising from the measurement error in the data was that industry mean responses from lines of business with only one or two respondents tended to be disproportionately located near the extremes of the distribution of mean responses to any given question. Most conclusions based on the full sample of 130 lines, and virtually all those emphasized in this paper, were replicated in the smaller sample of 75 lines that had more than two respondents.
23. We designed the questionnaire to ensure that cross-question comparisons would arise naturally in the minds of the respondents. The items were arranged in blocks, with each item in a block rated on the same semantic scale.
24. This view of the efficacy of sales and service efforts is consistent with the emphasis given to investment in "cospecialized assets" as a means of appropriation in David J. Teece, "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy," *Research Policy*, vol. 15 (December 1986), pp. 285-305.
25. See Ignatius Horstmann, Glenn M. MacDonald, and Alan Slivinski, "Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent," *Journal of Political Economy*, vol. 93 (October 1985), pp. 837-58, for a theoretical treatment of the issue.
26. To preserve confidentiality, we do not identify any industry in which there was only one response. Hereafter, we refer to such cases as singletons.
27. The same pattern appears when the survey data are aggregated up to the level (roughly two and one-half digit) at which the National Science Foundation reports detailed data on the extent and composition of research and development expenditures. Of the twenty-five industries into which the manufacturing sector is divided, only industrial chemicals, drugs, and petroleum refining rated process patents higher than four points, and only industrial chemicals and drugs rated product patents higher than five.
28. Our results were reinforced by Edwin Mansfield's finding that among the twelve broadly defined industries he studied only in the drug industry were patents considered essential to developing and marketing most inventions. Chemicals was the only other industry that considered patents essential for as many as 30 percent of inventions. See "Patents and Innovation: An Empirical Study," *Management Science*, vol. 32 (February 1986), pp. 173-81.
29. Yet another motive discussed in the literature is to gain strategic advantage in negotiation. In the semiconductor industry, for instance, the cumulative nature of the technology makes it difficult to participate legally without access to the patents of numerous firms. In consequence, there is widespread cross-licensing. Established firms, however, rarely license a new entrant until it has established a significant position in the market. As a defense against infringement suits, a prudent new entrant will establish a patent portfolio of its own, thus compelling established firms to negotiate cross-license agreements. See Eric von Hippel, "Appropriability of Innovation Benefit as a Predictor of the Source of Innovation," *Research Policy*, vol. 11 (January 1982), pp. 95-115; and Levin, "Semiconductor Industry," pp. 80-81.
30. Simple correlation coefficients were calculated using the individual respondent and industry mean responses as the units of observation. Correlations among industry means for the entire sample of 130 lines of business were qualitatively similar to those obtained when the sample was restricted to those with more than two responses. These and other correlation matrices discussed in this paper are available from the authors upon request.
31. The results reported in table 3 are based on a principal components analysis undertaken at the level of individual responses. An analysis at the level of industry mean responses produced similar results.
32. Additional evidence of the internal consistency of the survey results was provided by the pattern of negative correlation between responses concerning limitations on patent effectiveness and responses concerning the effectiveness of patents. Using either individual respondents or industry means as the unit of observation, all such correlation coefficients were negative except in the case of compulsory licensing. Most correlations were significant at the 0.1 level.

33. For example, P.L. 98-622, passed in 1984, modified the previous requirement that each coinventor listed in a patent application also had to be a coinventor on every claim of the patent. The new law allows inventors to apply jointly, even though they may not have physically worked together, made the same level of contribution, or contributed individually to the subject matter of each claim. For a thorough discussion, see Patrick Kelley, "Recent Changes in the Patent Law Which Affect Invention and the Ownership of Patents," unpublished manuscript (1985).
34. F. M. Scherer, *The Economic Effects of Compulsory Patent Licensing* (New York University, Graduate School of Business Administration, 1977).
35. Richard R. Nelson and Sidney G. Winter, "The Schumpeterian Tradeoff Revisited," *American Economic Review*, vol. 72 (March 1982), pp. 114-32; Michael Spence, "Cost Reduction, Competition, and Industry Performance," *Econometrica*, vol. 52 (January 1984), pp. 101-21; and Levin and Reiss, "Tests of a Schumpeterian Model," and "Demand-Creating and Cost-Reducing R&D."
36. Wesley Cohen and Daniel Levinthal have studied the incentives to engage in R&D that is directed toward developing absorptive capacity, the ability to make use of technology developed by others. See "Innovation and Learning: The Two Faces of R&D" (Carnegie Mellon University, Department of Social and Decision Sciences, March 1987).
37. The correlations between the effectiveness of particular learning mechanisms and the effectiveness of alternative methods of appropriation are interesting and internally consistent. In particular, when patent protection is effective, learning tends to take place primarily through licensing and patent disclosures. The effectiveness of patents is essentially uncorrelated with the effectiveness of interpersonal channels of learning and of independent R&D, and it is negatively correlated with the effectiveness of reverse engineering.
38. With three clusters the ratio of variance among clusters to variance within clusters was low, but attempts to find more than three clusters were thwarted by the persistent appearance of clusters containing only one or two lines of business.
39. See Edwin Mansfield, *Industrial Research and Technological Innovation: An Econometric Analysis* (Norton, 1968), for evidence on the time-cost trade-off.
40. Qualitatively identical results and interpretations are obtained from frequency distributions of individual responses and from the distribution of industry means.
41. Mansfield, Schwartz, and Wagner, "Imitation Costs and Patents."
42. The ranges are shown in the headings of table 8. The fifth and sixth column headings are not readily quantified. To permit the comparison discussed in the text, we assigned these categories the values of 112.5 percent and 137.5 percent, respectively, thereby maintaining a constant spacing of 25 percentage points between each pair of categories.
43. Our results on the time required to duplicate a rival's new products or processes were also roughly consistent with recent findings of Edwin Mansfield. In all but one of the ten industries he surveyed, the median respondent indicated that six to twelve months usually elapsed before the nature and operation of a new product were known to a firm's rivals. Effective duplication, as we have defined it, should take as long or longer, and table 9 shows that it typically does. The median and modal industries require one to three years to duplicate a major innovation or a typical patented innovation. A typical unpatented innovation, however, is more often duplicated within six to twelve months. See "How Rapidly Does New Industrial Technology Leak Out?" *Journal of Industrial Economics*, vol. 34 (December 1985), pp. 217-24.
44. The ratio of company-financed R&D to sales (R&D intensity) varies considerably among industries defined at the FTC line-of-business level of aggregation. In the 1976 data used by Levin, Cohen, and Mowery, R&D intensity ranged from 0.08 percent to 8.5 percent; both the mean and standard deviation were 1.7 percent. See "R&D Appropriability, Opportunity, and Market Structure."
45. Respondents were asked to identify, on a seven-point Likert scale ranging from "very slowly" to "very rapidly," the rate at which new processes and products had been introduced in their industries since 1970. Industry mean responses were highly correlated with total factor productivity growth, and the plausibility of the responses was reinforced by the identity of the highest and lowest industries in the sample. Excluding singletons, the lines of business reporting the slowest rates of product introduction were concrete, cement, boiler shops, milk, gypsum, primary copper, grain mill products, and sawmills. Those reporting the most rapid rates of product introduction were electrical equipment for internal combustion engines, radio and TV sets, computers, semiconductors, communications equipment, photographic equipment and supplies, engineering and scientific instruments, and guided missiles. Levin, Cohen, and Mowery used as a dependent variable the average of each industry's reported rates of process and product introduction.

46. Cohen, Levin, and Mowery, "Firm Size and R&D Intensity."
47. All coefficients in the R&D and innovation-rate equations were statistically significant at the .01 level.
48. To represent appropriability conditions, Levin, Cohen, and Mowery used two survey-based measures: the maximum of the mean scores an industry's respondents assigned to the effectiveness of the six methods of appropriation and the time required to duplicate effectively a patented major product innovation. To represent opportunity conditions, they used a measure of an industry's closeness to science as well as variables summarizing the importance of four other external sources of knowledge for an industry's technological advance: material suppliers, equipment suppliers, users of the industry's products, and government agencies and research labs.
49. It may seem anomalous that the effectiveness of appropriation was positively related to innovation but not to R&D, but the relationship was observed at the level of the industry. Better appropriability may discourage R&D directed toward imitation to an extent that more than compensates for its stimulus to innovative R&D. Such a reallocation of effort would be entirely consistent with the observed positive relationship between appropriability and the rate of innovation.
50. Levin and Reiss have used survey-based measures of appropriability and opportunity in a simultaneous equation model of R&D spending and market structure that builds on their work in "Tests of a Schumpeterian Model." Cohen and Levinthal use survey based variables in their work on R&D as investment in absorptive capacity; see "Innovation and Learning." Iain Cockburn and Zvi Griliches are studying the usefulness of our survey measures in estimating the value of patents from stock market data; see "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents," *American Economic Review* (forthcoming, May 1988). Meryl Finkel, "Overseas Research and Development by U.S. Multinationals: Ownership Structure Decisions" (Ph.D. dissertation, Harvard University, 1986), explored the effect of our appropriability variables on the investment decisions of multinational corporations. Franco Malerba is using the survey data to explain inter-industry differences in the extent and effectiveness of learning mechanisms.
51. For a calculation of the impact of the Drug Price Competition and Patent Term Restoration Act of 1984, see Henry Grabowski and John Vernon, "Longer Patents for Lower Imitation Barriers: The 1984 Drug Act," *American Economic Review*, vol. 76 (May 1986, *Papers and Proceedings*, 1985), pp. 195-98.
52. We focus on this particular consequence of the proposed legislation and set aside two major considerations regarding its merits in its present form. First, without amendment the legislation is likely to undercut severely the per se treatment of price fixing. Second, it might be more appropriate to consider eliminating per se treatment of all tying arrangements rather than just those involving intellectual property. On this point, see the concurring opinion in *Jefferson Parish v. Hyde*, 466 U.S. 2 (1984).
53. See E. W. Kitch, "Patents: Monopolies or Property Rights," in John Palmer, ed., *Research in Law and Economics: The Economics of Patents and Copyrights*, vol. 8 (JAI Press, 1986), pp. 31-47; and the associated commentary of F. M. Scherer, p. 51. *Digidyne Corp. v. Data General Corp.*, 743 F. 2d 1336 (9th Cir. 1984) is an example of an application of per se doctrine in a context where the intellectual property (software) does not convincingly convey market power.
54. The trade representative may at his discretion escape this requirement by finding that such an investigation would not be in the national interest.

Comments and Discussion

1. Christopher T. Taylor and Z. A. Silberston, *The Economic Impact of the Patent System: A Study of the British Experience* (Cambridge University Press, 1973); and Edwin Mansfield, Mark Schwartz, and Samuel Wagner, "Imitation Costs and Patents: An Empirical Study," *Economic Journal*, vol. 91 (December 1981), pp. 907-18.
 2. Partha Dasgupta and Joseph E. Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, vol. II (Spring 1980), pp. 1-28.
 3. Drew Fudenberg and others, "Preemption, Leapfrogging and Competition in Patent Races," *European Economic Review*, vol. 22 (June 1983), pp. 3-31.
 4. See William D. Nordhaus, *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (MIT Press, 1969).
1. Y. Grunfeld and Zvi Griliches, "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, vol. 42 (February 1960), pp. 1-13.
 3. Richard C. Levin, Wesley M. Cohen, and David C. Mowery, "R&D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review*, vol. 75 (May 1985, *Papers and Proceedings*,

1984), p. 23.

4. Iain Cockburn and Zvi Griliches, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents," working paper 2465 (National Bureau of Economic Research, December 1987).