

ISSN 1045-6333

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

ANTITRUST

Louis Kaplow
Carl Shapiro

Discussion Paper No. 575

01/2007

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

The Social Science Research Network Electronic Paper Collection:
http://papers.ssrn.com/abstract_id=#####

Antitrust*

Louis Kaplow[†] and Carl Shapiro[‡]

Abstract

This is a survey of the economic principles that underlie antitrust law and how those principles relate to competition policy. We address four core subject areas: market power, collusion, mergers between competitors, and monopolization. In each area, we select the most relevant portions of current economic knowledge and use that knowledge to critically assess central features of antitrust policy. Our objective is to foster the improvement of legal regimes and also to identify topics where further analytical and empirical exploration would be useful.

JEL Classes: K21, L12, L13, L40, L41, L42

© Louis Kaplow and Carl Shapiro. All rights reserved.

* We are grateful to Jonathan Baker for extensive and very valuable comments, Stephanie Gabor, Jeffrey Harris, Christopher Lanese, Bradley Love, Stephen Mohr, Andrew Oldham, Peter Peremiczki, Michael Sabin, and Kevin Terrazas for research assistance, and the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support.

[†] Harvard University and National Bureau of Economic Research

[‡] University of California at Berkeley

Antitrust*

Louis Kaplow[†] and Carl Shapiro[‡]

9 January 2007

Abstract

This is a survey of the economic principles that underlie antitrust law and how those principles relate to competition policy. We address four core subject areas: market power, collusion, mergers between competitors, and monopolization. In each area, we select the most relevant portions of current economic knowledge and use that knowledge to critically assess central features of antitrust policy. Our objective is to foster the improvement of legal regimes and also to identify topics where further analytical and empirical exploration would be useful.

Key Words: antitrust, competition policy, monopoly, market power, market definition, oligopoly, collusion, cartels, price fixing, facilitating practices, mergers, horizontal mergers, unilateral effects, monopolization, exclusionary practices, predatory pricing, exclusive dealing

JEL Keys: K21, L12, L13, L40, L41, L42

Forthcoming in *Handbook of Law and Economics*, A. Mitchell Polinsky & Steven Shavell, eds., Elsevier, 2007.

* We are grateful to Jonathan Baker for extensive and very valuable comments, Stephanie Gabor, Jeffrey Harris, Christopher Lanese, Bradley Love, Stephen Mohr, Andrew Oldham, Peter Peremiczki, Michael Sabin, and Kevin Terrazas for research assistance, and the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support.

[†] Harvard University and National Bureau of Economic Research

[‡] University of California at Berkeley

Table of Contents

I.	Introduction.....	1
II.	Market Power.....	2
A.	Definition of Market Power	3
B.	Single-Firm Pricing Model Accounting for Rivals	4
C.	Multiple-Firm Models	7
1.	Cournot Model with Homogeneous Products.....	7
2.	Bertrand Model with Differentiated Products.....	9
3.	Other Game-Theoretic Models and Collusion.....	11
D.	Means of Inferring Market Power	11
1.	Price-Cost Margin.....	12
2.	Firm's Elasticity of Demand.....	14
3.	Conduct.....	18
E.	Market Power in Antitrust Law.....	20
III.	Collusion	22
A.	Economic and Legal Approaches: An Introduction.....	23
1.	Economic Approach.....	23
2.	Legal Approach.....	24
B.	Oligopoly Theory	26
1.	Elements of Successful Collusion.....	26
2.	Repeated Oligopoly Games and the Folk Theorem.....	27
3.	Role of Communications	29

C.	Industry Conditions Bearing on the Likelihood of Collusive Outcomes	31
1.	Limited Growth for Defecting Firm	31
2.	Imperfect Detection	32
3.	Credibility of Punishment	33
4.	Market Structure	35
5.	Product Differentiation	39
6.	Capacity Constraints, Excess Capacity, and Investment in Capacity	40
7.	Market Dynamics	42
D.	Agreements under Antitrust Law	44
1.	On the Meaning of Agreement	44
2.	Agreement, Economics of Collusion, and Communications	47
E.	Other Horizontal Arrangements	51
1.	Facilitating Practices	52
2.	Rule of Reason	54
F.	Antitrust Enforcement	58
1.	Impact of Antitrust Enforcement on Oligopolistic Behavior	58
2.	Determinants of the Effectiveness of Antitrust Enforcement	59
IV.	Horizontal Mergers	59
A.	Oligopoly Theory and Unilateral Competitive Effects	61
1.	Cournot Model with Homogeneous Products	61
2.	Bertrand Model with Differentiated Products	65
3.	Bidding Models	70
B.	Oligopoly Theory and Coordinated Effects	71
C.	Empirical Evidence on the Effects of Horizontal Mergers	74
1.	Stock Market Prices	75
2.	Accounting Measures of Firm Performance	76
3.	Case Studies	76

D.	Antitrust Law on Horizontal Mergers	78
1.	Background and Procedure	78
2.	Anticompetitive Effects	81
3.	Efficiencies	83
E.	Market Analysis Under the Horizontal Merger Guidelines	89
1.	Market Definition: General Approach and Product Market Definition.....	90
2.	Geographic Market Definition.....	95
3.	Rivals' Supply Response	97
F.	Predicting the Effects of Mergers.....	98
1.	Direct Evidence From Natural Experiments.....	98
2.	Merger Simulation	99
V.	Monopolization	100
A.	Monopoly Power: Economic Approach.....	101
1.	Rationale for Monopoly Power Requirement.....	101
2.	Application to Challenged Practices.....	103
B.	Legal Approach to Monopolization	106
1.	Monopoly Power.....	106
2.	Exclusionary Practices	110
C.	Predatory Pricing	113
1.	Economic Theory.....	113
2.	Empirical Evidence.....	115
3.	Legal Test.....	116
D.	Exclusive Dealing.....	121
1.	Anticompetitive Effects	121
2.	Efficiencies	127
3.	Legal Test.....	128
VI.	Conclusion	130
VII.	References.....	131

I. Introduction

In this chapter we survey the economic principles that underlie antitrust law and use these principles to illuminate the central challenges in formulating and applying competition policy. Our twin goals are to inform readers about the current state of knowledge in economics that is most relevant for understanding antitrust law and policy and to critically appraise prevailing legal principles in light of current economic analysis.

Since the passage of the Sherman Act in 1890, antitrust law has always revolved around the core economic concepts of competition and market power. For over a century, it has been illegal in the United States for competitors to enter into price-fixing cartels and related schemes and for a monopolist to use its market power to stifle competition. In interpreting the antitrust statutes, which speak in very general terms, U.S. courts have always paid attention to economics. Yet the role of economics in shaping antitrust law has evolved greatly, especially over the past few decades. The growing influence of economics on antitrust law can be traced in part to the Chicago School, which, starting in the 1950s, launched a powerful attack on many antitrust rules and case outcomes that seemed to lack solid economic underpinnings. But the growing influence of economics on antitrust law also has resulted from substantial theoretical and empirical advances in industrial organization economics over the period since then. With a lag, often spanning a couple of decades, economic knowledge shapes antitrust law. It is our hope in this essay both to sharpen economists' research agendas by identifying open questions and difficulties in applying economics to antitrust law, and also to accelerate the dissemination of economic knowledge into antitrust policy.

Antitrust economics is a broad area, overlapping to a great extent with the field of industrial organization. We do not offer a comprehensive examination of the areas within industrial organization economics that are relevant for antitrust law. That task is far too daunting for a single survey and is already accomplished in the form of the three-volume *Handbook of Industrial Organization* (1989a, 1989b, 2007).¹ Instead, we focus our attention on four core economic topics in antitrust: the concept of market power (Section II), the forces that facilitate or impede efforts by competitors to engage in collusion (Section III), the effects of mergers between competitors (Section IV), and some basic forms of single-firm conduct that can constitute illegal monopolization, namely predatory pricing and exclusive dealing (Section V).² In each case, we attempt to select from the broad base of models and approaches the ones that seem most helpful in formulating a workable competition policy. Furthermore, we use this analysis to scrutinize the corresponding features of antitrust law, in some cases providing a

¹ Schmalensee and Willig (1989a, 1989b) and Armstrong and Porter (2007).

² Since the field of antitrust economics and law is far too large to cover in one chapter, we are forced to omit some topics that are very important in practice and have themselves been subject to extensive study, including joint ventures (touched on briefly in subsection III.E.2), vertical mergers, bundling and tying, vertical intrabrand restraints, the intersection of antitrust law and intellectual property law, and most features of enforcement policy and administration, including international dimensions.

firmer rationalization for current policy and in others identifying important divergences.³ For reasons of concreteness and of our own expertise, we focus on antitrust law in the United States, but we also emphasize central features that are pertinent to competition policy elsewhere and frequently relate our discussion to the prevailing regime in the European Union.⁴

II. Market Power

The concept of market power is fundamental to antitrust economics and to the law. Except for conduct subject to *per se* treatment, antitrust violations typically require the government or a private plaintiff to show that the defendant created, enhanced, or extended in time its market power. Although the requisite degree of existing or increased market power varies by context, the nature of the inquiry is, for the most part, qualitatively the same.

It is important to emphasize at the outset that the mere possession of market power is not a violation of antitrust law in the United States. Rather, the inquiry into market power is usually a threshold question; if sufficient market power is established, it is then asked whether the conduct in question—say, a horizontal merger or an alleged act of monopolization—constitutes an antitrust violation. If sufficient market power is not demonstrated, the inquiry terminates with a victory for the defendant.

Here, we begin our treatment of antitrust law and economics with a discussion of the basic economic concept of market power and its measurement. We first define market power, emphasizing that, as a technical matter, market power is a question of degree. Then we explore the factors that determine the extent of market power, first when exercised by a single firm and then in the case in which multiple firms interact. We also consider various methods of inferring market power in practice and offer some further remarks about the relationship between the concept of market power as understood by economists and as employed in antitrust law.⁵ Further elaboration appears in Sections IV and V on horizontal mergers and monopolization, respectively.

³ There are a number of books that have overlapping purposes, including Bork (1978), Hylton (2003), Posner (2001), and Whinston (2006), the latter being closest to the present essay in the weight given to formal economics.

⁴ As implied by the discussion in the text, our references to the law are primarily meant to make concrete the application of economic principles (and secondarily to offer specific illustrations) rather than to provide detailed, definitive treatments. On U.S. law, the interested reader should consult the extensive treatise *Antitrust Law* by Areeda and Hovenkamp, many volumes of which are cited throughout this essay. On the law in the European Union, see, for example, Bellamy and Child (2001), Dabbath (2004), and Walle de Ghelcke and van Gerven (2004). A wide range of additional information, including formal policy statements and enforcement statistics, are now available on the Internet. Helpful links are: Antitrust Division, Department of Justice: <http://www.usdoj.gov/atr/index.html>; Bureau of Competition, Federal Trade Commission: <http://www.ftc.gov/ftc/antitrust.htm>; European Union, DG Competition: http://ec.europa.eu/comm/competition/index_en.html; Antitrust Section of the American Bar Association, <http://www.abanet.org/antitrust/home.html>.

⁵ Prior discussions of the general relationship between the economic conception of market power and its use in antitrust law include Areeda, Kaplow, and Edlin (2004, pp. 483-499), Kaplow (1982), and Landes and Posner (1981). For a recent overview, see American Bar Association, Section of Antitrust Law (2005).

A. Definition of Market Power

Microeconomics textbooks distinguish between a price-taking firm and a firm with some power over price, that is, with some market power. This distinction relates to the demand curve facing the firm in question. Introducing our standard notation for a single firm selling a single product, we write P for the price the firm receives for its product, X for the firm's output, and $X(P)$ for the demand curve the firm perceives that it is facing, with $X'(P) \leq 0$.⁶ When convenient, we will use the inverse demand curve, $P(X)$. A price-taking firm has no control over price:

$P(X) = P$ regardless of X , over some relevant range of the firm's output. In contrast, a firm with power over price can cause price to rise or fall by decreasing or increasing its output: $P'(X) < 0$ in the relevant range. We say that a firm has "technical market power" if it faces a downward sloping (rather than horizontal) demand curve.

In practice almost all firms have some degree of technical market power. Although the notion of a perfectly competitive market is extremely useful as a theoretical construct, most real-world markets depart at least somewhat from this ideal. An important reason for this phenomenon is that marginal cost is often below average cost, most notably for products with high fixed costs and few or no capacity constraints, such as computer software, books, music, and movies. In such cases, price must exceed marginal cost for firms to remain viable in the long run.⁷ Although in theory society could mandate that all prices equal marginal cost and provide subsidies where appropriate, this degree of regulation is generally regarded to be infeasible, and in most industries any attempts to do so are believed to be inferior to reliance upon decentralized market interactions. Antitrust law, as explained in the introduction, has the more modest but, it is hoped, achievable objective of enforcing competition to the extent feasible. Given the near ubiquity of some degree of technical market power, the impossibility of eliminating it entirely, and the inevitable costs of antitrust intervention, the mere fact that a firm enjoys some technical market power is not very informative or useful in antitrust law.

Nonetheless, the technical, textbook notion of market power has the considerable advantage that it is amenable to precise measurement, which makes it possible to identify practices that enhance a firm's power to a substantial degree. The standard measure of a firm's technical market power is based on the difference between the price the firm charges and the firm's marginal cost. In the standard theory of monopoly pricing, a firm sets the price for its product to maximize profits. Profits are given by $\pi = PX(P) - C(X(P))$, where $C(X)$ is the firm's cost function.

Differentiating with respect to price, we get the standard expression governing pricing by a single-product firm,

⁶ For simplicity, unless we indicate otherwise, we assume throughout this chapter that each firm sells a single product. While this assumption is almost always false, in many cases it amounts to looking at a firm's operations product-by-product. Obviously, a multi-product firm might have market power with respect to one product but not others. When interactions between the different products sold by a multi-product firm are important, notably, when the firm sells a line of products that are substitutes or complements for each other, the analysis will need to be modified.

⁷ Edward Chamberlin (1933) and Joan Robinson (1933) are classic references for the idea that firms in markets with low entry barriers but differentiated products have technical market power.

$$\frac{P - MC}{P} = \frac{1}{|\varepsilon_F|} \quad (1)$$

where MC is the firm's marginal cost, $C'(X)$, and $\varepsilon_F \equiv \frac{dX}{dP} \frac{P}{X}$ is the elasticity of demand facing that firm, the "firm-specific elasticity of demand."⁸ The left-hand side of this expression is the Lerner Index, the percentage gap between price and marginal cost, which is a natural measure of a firm's technical market power:

$$m \equiv \frac{P - MC}{P}.$$

As noted earlier, some degree of technical market power is necessary for firms to cover their costs in the presence of economies of scale. For example, if costs are given by $C(X) = F + CX$, then profits are given by $\pi = PX - CX - F$ and the condition that profits are non-negative can be written as $m \geq F / PX$, that is, the Lerner Index must be at least as large as the ratio of the fixed costs, F , to the firm's revenues, $R \equiv PX$.

Before proceeding with our analysis, we note that, although anticompetitive harm can come in the form of reduced product quality, retarded innovation, or reduced product variety, our discussion will follow much of the economics literature and most antitrust analysis in focusing on consumer harm that comes in the form of higher prices. This limitation is not as serious as may first appear because higher prices can serve as a loose proxy for other forms of harm to consumers.

B. Single-Firm Pricing Model Accounting for Rivals

To aid understanding, we present a basic but flexible model showing how underlying supply and demand conditions determine the elasticity of demand facing a given firm. This model allows us to begin identifying the factors that govern the degree of technical market power enjoyed by a firm. We also note that this same model will prove very useful conceptually when we explore below the impact of various practices on price. Studying the effects of various practices on price requires some theory of how firms set their prices. The building block for these various theories is the basic model of price-setting by a single, profit-maximizing firm. In addition, as a matter of logic, one must begin with such a model before moving on to theories that involve strategic interactions among rival firms.

The standard model involves a dominant firm facing a competitive fringe.⁹ A profit-maximizing firm sets its price accounting for the responses it expects from its rivals and customers to the

⁸ Strictly speaking, the elasticity of demand facing the firm is endogenous, except in the special case of constant elasticity of demand, since it varies with price, an endogenous variable. All the usual formulas refer to the elasticity of demand at the equilibrium (profit-maximizing) price level.

⁹ For a recent textbook treatment of this model, see Carlton and Perloff (2005, pp. 110-119). Landes and Posner (1981) provide a nice exposition of this model in the antitrust context.

price it sets.¹⁰ This is a decision-theoretic model, not a game-theoretic model, so it does not make endogenous the behavior of the other firms in the market or of potential entrants. This is the primary sense in which the generality of the model is limited. The model also is limited because it assumes that all firms in the market produce the same, homogeneous product and do not engage in any price discrimination, although the core ideas underlying it extend to models of differentiated products.

The firm faces one or more rivals that sell the same, homogeneous product. When setting its price, P , the firm recognizes that rivals will likely respond to higher prices by producing more output. The combined output of the firm's rivals increases with price according to $Y(P)$, with $Y'(P) \geq 0$. Total (market) demand declines with price according to $Z(P)$, with $Z'(P) \leq 0$. If the firm in question sets the price P , then it will be able to sell an amount given by $X(P) \equiv Z(P) - Y(P)$. This is the largest quantity that the firm can sell without driving price below the level P that it selected; if the firm wants to sell more, it will have to lower its price. The firm's so-called "residual demand curve" is therefore given by $X(P)$.

If we differentiate the equation defining $X(P)$ with respect to P , and then multiply both sides by $-P/X$ to convert the left-hand side into elasticity form, we get

$$-\frac{P}{X} \frac{dX}{dP} = -\frac{P}{Z} \frac{dZ}{dP} + \frac{P}{Y} \frac{dY}{dP}.$$

Next, multiply and divide the dZ/dP term on the right-hand side by Z and the dY/dP term by Y . This gives

$$-\frac{P}{X} \frac{dX}{dP} = -\frac{P}{Z} \frac{dZ}{dP} \frac{Z}{X} + \frac{P}{Y} \frac{dY}{dP} \frac{Y}{X}.$$

Define the market share of the firm being studied by $S = X/Z$. The corresponding market share of the rivals is $1 - S = Y/Z$. Replacing Z/X by $1/S$ and Y/X by $(1 - S)/S$ in the expression above gives

$$-\frac{P}{X} \frac{dX}{dP} = -\frac{P}{Z} \frac{dZ}{dP} \frac{1}{S} + \frac{P}{Y} \frac{dY}{dP} \frac{(1 - S)}{S}.$$

¹⁰ As with the standard theory of pure monopoly pricing as taught in microeconomics textbooks, the results of this model are unchanged if we model the firm as choosing its output level, with price adjusting to clear the market.

Call the elasticity of supply of the rivals $\varepsilon_R \equiv \frac{P}{Y} \frac{dY}{dP}$, and the absolute value of the elasticity of the underlying market demand curve $|\varepsilon_D| \equiv -\frac{P}{Z} \frac{dZ}{dP}$. The absolute value of the elasticity of demand facing the firm, $|\varepsilon_F| \equiv -\frac{P}{X} \frac{dX}{dP}$, is therefore given by

$$|\varepsilon_F| = \frac{|\varepsilon_D| + (1 - S)\varepsilon_R}{S}. \quad (2)$$

This equation captures the central lesson from this model: the absolute value of the elasticity of demand facing a single firm, given the supply curves of its price-taking rivals and the demand curve of the buyers in its market, is governed by three variables: (1) the underlying elasticity of demand for the product, $|\varepsilon_D|$, which as noted is frequently called the market elasticity of demand; (2) the elasticity of supply of the firm's rivals, ε_R ; and (3) the firm's market share, S . The magnitude of the firm-specific elasticity of demand is larger, the larger are the magnitude of the market elasticity of demand and the elasticity of supply of the firm's rivals and the smaller is the firm's market share. Intuitively, market share is relevant for two reasons: the smaller the firm's share, the greater the share of its rivals and thus the greater is the absolute magnitude of their supply response to a price increase for a given supply elasticity, ε_R ; and the smaller the firm's share, the smaller is its share of the increase in industry profits due to a given sacrifice in its own sales.¹¹

One polar case in this basic model is that of the traditional monopolist. With no rivals, $S = 1$, so the elasticity of demand facing the firm is just the market elasticity of demand. With rivals, however, the magnitude of the firm-specific elasticity of demand is larger than that of the market elasticity of demand. The other polar case is that of the firm from the theory of perfectly competitive markets. As the firm's share of the market approaches zero, the magnitude of the firm-specific elasticity of demand becomes infinite, that is, the firm is a price-taker.

We can directly translate the firm-specific elasticity of demand given by expression (2) into the profit-maximizing price. As indicated in expression (1), profit maximization involves setting price so that the firm's gross margin, m , equals the inverse of the magnitude of the firm's elasticity of demand. If there are no rivals, $S = 1$ and this relationship simplifies to the standard monopoly formula, $m = 1/|\varepsilon_D|$. For a firm with a tiny market share, ε_F is enormous, so $m \approx 0$, that is, price nearly equals marginal cost. For intermediate cases, as noted, in this model a large market elasticity of demand, $|\varepsilon_D|$, a high elasticity of rival supply, ε_R , and a small market share,

¹¹ It should be noted that statements about the effect of market share must be interpreted carefully. Thus, an outward shift in the supply curve of the rivals, which lowers the firm's market share at any given price, will raise the elasticity of demand facing that firm at any given price. However, more broadly, the firm's market share is endogenous because it depends on the price the firm chooses.

S , all lead to a large firm-specific elasticity of demand facing the price leader, $|\varepsilon_F|$, which in turn implies a small margin.

This model provides a guide for studying the types of conduct that may enhance a firm's technical market power and thus allow that firm profitably to raise its price. Generically, such conduct will be that which reduces the value of the right side of expression (2): conduct that makes substitute products less attractive, that causes rivals to reduce their supply, and that raises the firm's market share (through the two former means or otherwise). Later we consider how certain types of conduct having these effects should be scrutinized under antitrust law.

This model is quite broad when one undertakes appropriate interpretations and extensions. For example, issues relating to substitute products bear on the market elasticity of demand, as will be noted below. Additionally, one can account for entry by reflecting it in the rival supply elasticity. One particular variant of the model involves infinitely elastic rival supply, perhaps due to entry, at some fixed "limit" price.

C. Multiple-Firm Models

The model in Section B took the behavior of all but one firm as exogenous. In this section, we consider game-theoretic models that make predictions regarding the degree of market power exercised by interacting firms. First we consider two standard, static, noncooperative models: Cournot's model of oligopoly, for the case with homogeneous products, and Bertrand's model, for the case with differentiated products. Then we consider briefly the possibility of repeated games and the impact of collusive behavior on market power.¹²

1. Cournot Model with Homogeneous Products

The Cournot (1838) model of oligopoly with homogeneous products is similar to the single-firm pricing model in that it identifies how certain observable characteristics of the market determine the degree of a firm's market power, that is, the percentage markup above marginal cost that the firm charges. The Cournot model goes further, however, by providing predictions about how market structure affects the equilibrium price, predictions that will be important for seeing how certain commercial practices and mergers affect price. Specifically, the model predicts that firms with lower costs will have higher market shares and higher markups. The model is frequently employed in markets with relatively homogeneous products, especially if firms pick their output or capacity levels, after which prices are determined such that the resulting supply equals demand.¹³ However, one should bear in mind that the Cournot equilibrium is the Nash equilibrium in a one-shot game. As we discuss at length in Section III, many different outcomes can arise as equilibria in a repeated oligopoly game, even if the stage game played each period

¹² There is an enormous literature on oligopoly theory, which we do not attempt to cover systematically. See, for example, Shapiro (1989), Tirole (1988), and Vives (2001). We discuss models of repeated oligopoly at greater length in Section III on collusion.

¹³ Kreps and Scheinkman (1983) use a particular rationing rule to show that capacity choices followed by pricing competition can replicate the Cournot equilibrium.

involves quantity-setting à la Cournot. In antitrust applications, it is generally desirable to test robustness of results to alternative solution concepts as well as to test empirically the predictions of any oligopoly model that is employed.

In a Cournot equilibrium, a single firm's reaction curve is derived as a special case of the basic model of single-firm pricing: the rivals' outputs are all taken to be fixed, so the rival supply elasticity is zero. As we now show, the elasticity of demand facing a single firm is equal to the market elasticity of demand divided by that firm's market share. However, the Cournot model goes beyond the single-firm pricing model because it involves finding the equilibrium in a game among multiple firms.

Suppose that there are N firms, with each firm i choosing its output X_i simultaneously. The Cournot equilibrium is a Nash equilibrium in these quantities. Total output is $X \equiv X_1 + \dots + X_N$. Industry or market (inverse) demand is given by $P = P(X)$. Given the output of the other firms, firm i chooses its output to maximize its own profits, $\pi_i = P(X)X_i - C_i(X_i)$. The first-order condition for this firm is $P(X) + X_i P'(X) - C_i'(X_i) = 0$. This can be written as

$$\frac{P - MC_i}{P} = \frac{S_i}{|\varepsilon_D|} \quad (3)$$

where $S_i \equiv X_i / X$ is firm i 's market share, and ε_D , as before, is the market elasticity of demand.

To explore this result, consider the special case in which each firm i has constant marginal cost MC_i . Adding up the first-order conditions for all of the firms gives $NP(X) + XP'(X) = \sum_i MC_i$,

which tells us that total output and hence the equilibrium price depend only upon the sum of the firms' marginal costs. Moreover, the markup equation tells us that lower-cost firms have higher market shares and enjoy more technical market power. At the same time, the larger is the market elasticity of demand for this homogeneous product, the smaller is the market power enjoyed by each firm and the lower are the margins at all firms. Here we see a recurrent theme in antitrust: a lower-cost firm may well enjoy some technical market power and capture a large share of the market, but this is not necessarily inefficient. Indeed, with constant marginal costs, full productive efficiency would call for the firm with the lowest marginal costs to serve the entire market.

The Cournot model also predicts that total output will be less than would be efficient because none of the firms produces up to the point at which marginal cost equals price; they all have some degree of market power. In the special case with constant and equal marginal costs, each firm has a market share of $1/N$, and the model predicts that each enjoys technical market power according to the resulting equation $(P - MC)/P = 1/N |\varepsilon_D|$. In this simple sense, more firms leads to greater competition and lower prices. However, this model is clearly incomplete for antitrust purposes: presumably, there are fixed costs to be covered (which is why there is a fixed

number of firms in the first place), so adding more firms is not costless.¹⁴ This type of analysis will be directly relevant when we consider horizontal mergers, which remove an independent competitor but may also lead to efficiencies of various types.

One of the attractive theoretical features of the Cournot model is that it generates an elegant formula for the industry-wide average, output-weighted, price-cost margin, that is, the expression

$$PCM \equiv \sum_{i=1}^N S_i \frac{P - MC_i}{P}. \text{ Using equation (3), we get } PCM \equiv \sum_{i=1}^N S_i \frac{S_i}{|\varepsilon_D|} \text{ or}$$

$$PCM = \frac{1}{|\varepsilon_D|} \sum_{i=1}^N S_i^2 = \frac{H}{|\varepsilon_D|} \quad (4)$$

where $H \equiv \sum S_i^2$ is the Herfindahl-Hirschman Index (HHI) of market concentration that is commonly used in antitrust analysis, especially of horizontal mergers.

2. Bertrand Model with Differentiated Products

The Bertrand model with differentiated products is the other key static model of oligopoly used in antitrust. The Bertrand equilibrium is the Nash equilibrium in the game in which the firms simultaneously set their prices. With N firms selling differentiated products, we can write the demand for firm i 's product as $X_i = D_i(P_1, \dots, P_N)$. As usual, the profits of firm i are given by $\pi_i = P_i X_i - C_i(X_i)$. The Bertrand equilibrium is defined by the N equations $\partial \pi_i / \partial P_i = 0$.

Writing the elasticity of demand facing firm i as $\varepsilon_i \equiv \frac{\partial X_i}{\partial P_i} \frac{P_i}{X_i}$, firm i 's first-order condition is the usual markup equation,

$$\frac{P_i - MC_i}{P_i} = \frac{1}{|\varepsilon_i|}.$$

Actually solving for the Bertrand equilibrium can be difficult, depending on the functional form for the demand system and on the firms' cost functions. In general, however, we know that a firm faces highly elastic demand if its rivals offer very close substitutes, so the Bertrand theory predicts larger markups when the products offered by the various firms are more highly differentiated. In practice, notably, in the assessment of mergers, particular models of product differentiation are used, such as discrete choice models with random utilities, including logit and

¹⁴ In general, there is no reason to believe that the equilibrium number of firms in an oligopoly with free entry, that is, where equally efficient firms enter until further entry would drive profits below zero, is socially efficient. See, for example, Mankiw and Whinston (1986). This observation is relevant in assessing certain antitrust policies: if the equilibrium number of firms is "naturally" too small, then exclusionary conduct on the part of the incumbent oligopolists creates an additional social inefficiency. However, if the equilibrium number of firms is "naturally" excessive, different implications would follow.

nested logit models, or models with linear demand or constant elasticities, as we discuss further in Section IV on horizontal mergers.

Here we illustrate the operation of the Bertrand model by explicitly solving a simple, symmetric, two-firm model with constant marginal costs and linear demand. Write the demand curves as $X_1 = A - P_1 + \alpha P_2$ and $X_2 = A - P_2 + \alpha P_1$. Note that the parameter α measures the diversion ratio, that is, the fraction of sales lost by one firm, when it raises its price, that are captured by the other firm (assuming that the other firm's price is fixed). The diversion ratio, α , will be important when we study horizontal mergers below.¹⁵

Call the marginal costs per unit MC_1 and MC_2 , respectively, and assume that there are no fixed costs. Then we have $\pi_1 = (P_1 - MC_1)(A - P_1 + \alpha P_2)$. Differentiating with respect to P_1 and setting this equal to zero, we get firm 1's best-response curve, $P_1 = (A + \alpha P_2 + MC_1)/2$.

Assuming cost symmetry as well, $MC = MC_1 = MC_2$, in the symmetric Bertrand equilibrium we must have $P_1 = P_2 = P_B$ so we get $P_B = \frac{A + MC}{2 - \alpha}$.

We can compare the Bertrand equilibrium price to the price charged by a single firm controlling both products. Such a firm would set P to maximize $(P - MC)(A - P + \alpha P)$, which gives the monopoly price of $P_M = \frac{A + MC(1 - \alpha)}{2(1 - \alpha)}$. The percentage gap between the monopoly price and

the Bertrand price is given by $\frac{P_M - P_B}{P_B} = \frac{\alpha}{2(1 - \alpha)} \frac{P_B - MC}{P_B}$.¹⁶ This expression tells us that the

Bertrand equilibrium price is relatively close to the monopoly price when the two products are rather poor substitutes, that is, when the diversion ratio, α , is low.

This formula will be highly relevant when studying the effect on price of a merger between two suppliers of differentiated products. In that context, the formula measures the price increase associated with the merger, given the prices charged by other firms (and before accounting for efficiencies). The price increase will depend on the pre-merger margin, $\frac{P_B - MC}{P_B}$, and on the diversion ratio.

¹⁵ More generally, the diversion ratio from product i to substitute product j is defined as

$\alpha_{ji} = (dX_j / dP_i) / (-dX_i / dP_i)$. Converting this equation into elasticity form gives $\alpha_{ji} = \frac{\varepsilon_{ji}}{|\varepsilon_i|} \frac{X_j}{X_i}$, where

$\varepsilon_{ji} = \frac{dX_j}{dP_i} \frac{P_i}{X_j}$ is the cross-elasticity from product i to product j .

¹⁶ The details of these calculations are available at <http://faculty.haas.berkeley.edu/shapiro/unilateral.pdf>.

3. Other Game-Theoretic Models and Collusion

Both the Cournot and Bertrand models assume that firms engage in a one-shot noncooperative game. An extensive literature on repeated games explores the possibility that firms may do better for themselves, supporting what are more colloquially described as collusive outcomes, approaching or equaling the industry profit-maximizing price. As suggested by Stigler (1964) and refined in subsequent work, higher prices tend to be sustainable when cheating can be rapidly detected and effectively punished. For a general discussion of models of collusion, see Jacquemin and Slade (1989) and Shapiro (1989).

The possibility that firms can support alternative equilibria featuring higher prices is important to antitrust analysis. First, it suggests that market power may be higher than is otherwise apparent. Second and more important, the possibility of collusion affects the antitrust analysis of other business conduct. For example, a horizontal merger may have only a minor impact on price if the merging firms and their rivals are already colluding, but a far greater effect if the reduction in the number of competitors makes collusion easier to sustain. Also, some practices may facilitate collusion, in which case such practices themselves should potentially be subject to antitrust scrutiny. These possibilities are explored further in Section III on collusion and Section IV on horizontal mergers.

D. Means of Inferring Market Power

Assessing the extent of or increase in technical market power in a given situation is often a difficult undertaking. Based upon the foregoing analysis, one can identify a number of potential strategies whose usefulness varies greatly by context. The legal system has tended to rely primarily on a subset of these approaches, focusing mostly on market definition, as discussed below. In recent years, however, it has increasingly considered alternatives when it has perceived that credible economic evidence has been offered.¹⁷

Although somewhat crude, it is helpful to group means of inferring market power into three categories. First, since market power is technically defined by the extent of the price-cost margin, one can attempt to identify evidence that bears fairly directly on the size of this margin, or by measuring profits (which reflects the margin between price and *average* cost). Second, various models, such as the single-firm price-setting model in Section B, indicate that the extent of market power will be a function of the elasticity of demand, a firm's market share, and rivals' supply response. Accordingly, one can analyze information indicative of the magnitude of these factors. Third, one can make inferences from firm behavior, notably when observed actions would be irrational unless a certain degree of market power existed or was thereby conferred.

¹⁷ For example, the Supreme Court in *Federal Trade Commission v. Indiana Federation of Dentists*, 476 U.S. 447, 460-461 (1986) (quoting Areeda's *Antitrust Law* treatise) stated: "Since the purpose of the inquiries into market definition and market power is to determine whether an arrangement has the potential for genuine adverse effects on competition, 'proof of actual detrimental effects, such as a reduction of output,' can obviate the need for an inquiry into market power, which is but a 'surrogate for detrimental effects.'"

1. Price-Cost Margin

a) Direct Measurement

Observing the extent to which price is above marginal cost indicates the degree of technical market power. This direct approach is feasible if one can accurately measure price and some version of marginal cost, usually average incremental cost.¹⁸ Price is often easy to identify, although complications may arise when multiple products are sold together, making it difficult to determine the incremental revenue associated with the product in question. If different customers are charged different prices, it may be necessary to calculate the profit margins for sales to different customers (or at different points of time). Complexities also arise when some sales implicitly bundle other services, such as delivery, short-term financing, and customer support; in principle, these factors can be accommodated by redefining the product to include these services (and tracking the costs associated with these services). Marginal cost, by contrast, may be more difficult to measure, due both to difficulties in identifying which costs are variable (and over what time period) and to the presence of common costs that may be difficult to allocate appropriately. In part for this reason, the empirical industrial organization literature, as surveyed in Bresnahan (1989), often treats marginal cost as unobservable.

In some cases, approximate measures of price-cost margins may be sufficient and easy to produce, but as evidenced by disputes over cost in predatory pricing cases and in various regulatory contexts, direct measurement of any conception of cost can be difficult and contentious. In any event, as with all measures of technical market power, it is important to keep in mind the distinction between the extent of market power and whether particular conduct should give rise to antitrust liability. For example, as we have already noted, especially in industries in which marginal cost is below average cost and capacity constraints are not binding, nontrivial technical market power may be consistent with what are normally considered competitive industries.

b) Price Comparisons

Another fairly direct route to assessing the magnitude of price-cost margins, or at least to provide a lower-bound estimate, is to compare prices across markets. For example, if a firm sells its product for a substantially higher price in one region than in another (taking into account transportation and other cost differences), the price-cost margin in the high-price region should be at least as great as the (adjusted) price difference between the regions. This inference presumes, of course, that the price in the low-price region is at least equal to marginal cost. Note that this method can be understood as a special case of direct measurement. It is assumed that

¹⁸ We use the terms “marginal cost” and “average incremental cost” interchangeably. Both measure the extra cost per unit associated with increased output. Average incremental cost is a somewhat more accurate term, since one is often interested in increments that do not correspond to “one unit” of output. However, if one takes a flexible approach to what constitutes a “unit” of production, the two terms are exactly the same. In practice, average incremental cost is used to determine gross profit margins.

the low price is a proxy for (at least an upper bound on) marginal cost, and one then is measuring the price-cost margin directly.

The *Staples* merger case illustrates an application of this method.¹⁹ The government offered (and the court was convinced by) data indicating that prices were higher in regional markets in which fewer office supply superstores operated and that prices fell when new superstore chains entered. This was taken as powerful evidence that a merger of two of the existing three superstores would lead to price increases.

c) Price Discrimination

Price comparisons often involve a special case of price discrimination, wherein a given firm charges different prices to different consumers, contrary to the implicit assumption in the earlier analysis that each firm sets a single price for all of its customers. Accordingly, for essentially the same reason as that just given, the ability of a firm to engage in price discrimination implies the existence of market power. If one is prepared to assume that the firm is not pricing below marginal cost to any of its customers, and if one accounts for differences in the cost of serving different customers, the percentage difference between any high price it charges and the lowest price it charges for the same product can serve as a lower bound on the percentage markup associated with the higher price. For example, the substantial price discrimination in sales of pharmaceutical drugs on international markets shows that prices in the United States are very much above marginal cost.

The fact that price discrimination technically implies market power is important because price discrimination is widespread. Familiar examples include airline pricing, senior citizen and student discounts, and the mundane practice of restaurants charging steep price increments for alcoholic beverages (versus soft drinks) and high-end entrees that greatly exceed any differences in marginal cost. For business-to-business transactions, negotiations that typically generate price dispersion and price discrimination are quite common.

Once again, however, it is important to keep in mind that the technical existence of market power does not imply antitrust liability.²⁰ As is familiar, price discrimination generates greater seller profits yet may well be benign or even favorable on average for consumers. Moreover, the

¹⁹ Federal Trade Commission v. Staples, Inc., 970 F. Supp. 1066 (D.D.C. 1997). For further discussion, see subsection IV.F.1 in our discussion of horizontal mergers.

²⁰ Nor is it the case that price discrimination in itself implies antitrust liability, despite the existence of the Robinson-Patman Act that regulates particular sorts of price discrimination in certain contexts. As presently interpreted, price discrimination may be a violation in so-called primary-line cases, tantamount to predatory pricing, and in secondary-line cases, such as when manufacturers offer discounts (that are not cost justified) to large retailers that are not available to smaller buyers. Notably, the Act does not cover discriminatory prices to ultimate consumers (or to intermediaries that are not in competition with each other) that are nonpredatory. Nevertheless, it seems that defendants in antitrust litigation have been reluctant to rationalize challenged practices that analysts have suggested were means of price discrimination on such grounds, presumably fearing that such explanations would be to their detriment. Of course, one way this could be true is that the existence of some technical market power would thereby be conceded.

resulting profit margins are often necessary to cover fixed costs, as in models of monopolistic competition. If there are no barriers to entry so that the resulting margins merely provide a normal rate of return on capital, the presence of a gap between price and marginal cost is perfectly consistent with the conclusion that the market is behaving in a competitive fashion, given the presence of fixed costs and product differentiation. Furthermore, in our preceding example of multinational pharmaceutical companies, the margins provide the reward for costly and risky research and development to create and patent new drugs. The *ex post* market power is necessary to provide the quasi-rents that induce innovation (given that we rely on a patent system rather than a regime that gives direct rewards to innovators from the government fisc).

d) Persistent Profits

A somewhat different approach to establishing antitrust market power involves looking at a firm's profits, which amounts to comparing price to average (rather than marginal) cost. Under this approach, persistently above-normal profits indicate a high price-cost margin and thus the existence of technical market power. This method shares difficulties with any that rely on measures of cost. In particular, it is often very hard to measure the return on capital earned for a given product, or in a given market, especially for a firm that is engaged in many lines of business and has substantial costs that are common across products.²¹ Another problem with this approach is that the return on capital should, in principle, be adjusted for risk. Frequently, one is looking at a successful firm, perhaps one that has been highly profitable for many years following some initial innovation that, *ex ante*, may not have turned out as well.

In addition, average costs often differ from marginal costs. When average costs are higher, this approach may mask the existence of technical market power. In such circumstances, however, marginal-cost pricing may be unsustainable in any event; that is, although there may be technical market power, there may not be any way (short of intrusive regulation that is not contemplated) to improve the situation. When average cost is below marginal cost, profits can exist despite the absence of any markup. In such cases, entry might be expected. If profits are nevertheless persistent, there may exist entry barriers, a subject we discuss below.

2. Firm's Elasticity of Demand

In the single-firm pricing model, the price-cost margin (Lerner Index) equals the inverse of the (absolute value of the) firm's elasticity of demand, as indicated by expression (1). Furthermore, as described in expression (2), this elasticity depends on the market elasticity of demand, the firm's market share, and rivals' supply elasticity. In the Cournot, Bertrand, and other oligopoly models, many of the same factors bear on the extent of the price-cost margin and thus the degree of market power. Accordingly, another route to inferring market power is to consider the magnitude of these factors.

²¹ See, for example, Fisher and McGowan (1983).

a) Direct Measurement

One could attempt to measure the elasticity of demand facing the firm in question.²² A possible approach would be to estimate the market elasticity of demand and then make an adjustment based on the firm's market share. Alternatively, one might directly observe how the firm's sales have varied when it has changed its price. As a practical matter, both of these methods may be difficult to implement. However, they may nevertheless be more reliable than the alternatives.

b) Substitutes, Market Definition, and Market Share

In antitrust analysis, both by agencies (notably, in examining prospective horizontal mergers) and by the courts, the dominant method of gauging the extent of market power involves defining a so-called relevant market and examining the share of a firm or group of firms in that market. In defining product markets, the focus is on which products are sufficiently good demand substitutes for the product in question to be deemed in the same market. Likewise, in defining the extent of the geographic market, the question concerns the feasibility of substitution, for example, by asking how far patients would travel for hospitalization. Although we have discussed the economic analysis of market power at some length, the concept of market definition has not yet appeared directly. Hence it is useful to consider the relationship between the most common method used in antitrust law to assess market power and the implications of the foregoing economic analysis.

The connection is easiest to see by examining expression (2), which relates the firm-specific elasticity of demand to the market elasticity of demand, the firm's market share, and rivals' elasticity of substitution. Consider the case in which the firm produces a homogeneous product, has a high share of sales of that product, and faces a highly elastic market demand curve due to the existence of many close substitutes. The firm-specific elasticity of demand and thus the extent of technical market power is small even though the firm's market share is high in the narrowly defined market consisting only of the homogeneous product sold by the firm. One could redefine the "market" to include the close substitutes along with the homogeneous product sold by the firm. The market elasticity of demand in this broader market is presumably smaller, but since the firm's market share in this market is also necessarily lower, we would again conclude that the firm-specific demand elasticity is small and thus that the degree of technical market power is low as well.

Courts—and thus lawyers and government agencies—traditionally equate high market shares with a high degree of market power and low shares with a low degree of market power. This association is highly misleading if the market elasticity of demand is ignored, and likewise if rivals' elasticity of supply is not considered. In principle, as just explained, the paradigm based on market definition and market share takes the market elasticity of demand into account, indirectly, by defining broader markets—and thus producing lower market shares—when the elasticity is high. As should be apparent from the foregoing discussion, the standard antitrust

²² See, for example, Baker and Bresnahan (1988).

approach is more indirect than necessary and, due to this fact plus its dichotomous structure (substitutes are either in the market or not), will tend to produce needlessly noisy conclusions.²³ We discuss market definition at greater length in Section IV.E on horizontal mergers and subsection V.B.1 on monopolization.

Frequently, it is useful to decompose the elasticity of demand for a given product into various cross-elasticities of demand with other products. For example, if the price of soda rises, consumers will substitute to other drinks, including, perhaps, beer, juice, milk, and water. Naturally, the analysis in any given case will depend upon exactly how these various products are defined (soda could be broken into regular soda and diet soda, or colas and non-colas, etc.). But the underlying theory of demand does not vary with such definitions. To illustrate, suppose

that consumers allocate their total income of I across N distinct products, so $\sum_{i=1}^N P_i X_i = I$. To

study the elasticity of demand for product 1, suppose that P_1 rises and the other prices remain unchanged. Then we get $X_1 + P_1 \frac{dX_1}{dP_1} + \sum_{i=2}^N P_i \frac{dX_i}{dP_1} = 0$. Converting this to elasticity form gives

$-\frac{P_1}{X_1} \frac{dX_1}{dP_1} = 1 + \sum_{i=2}^N \frac{P_i}{X_i} \frac{dX_i}{dP_1} \frac{P_i X_i}{P_1 X_1}$. Defining the cross-elasticity between product i and product 1

as $\varepsilon_{i1} = \frac{dX_i}{dP_1} \frac{P_1}{X_i}$, and the revenues associated with product i as $R_i = P_i X_i$, this can be written as

$$|\varepsilon_{11}| = 1 + \sum_{i=2}^N \varepsilon_{i1} \frac{R_i}{R_1}. \quad (5)$$

In words, the (absolute value of the) elasticity of demand for product 1 is equal to one plus the sum of the cross-elasticities of all the other products with product 1, with each cross-elasticity weighted by the associated product's revenues relative to those of product 1. If we define each product's share of expenditures as $s_i = R_i / I$, then expression (5) can be written as

$|\varepsilon_{11}| = 1 + \frac{1}{s_1} \sum_{i=2}^N s_i \varepsilon_{i1}$, so the cross-elasticity with each rival product is weighted by its share of industry revenues.²⁴

This decomposition of the market elasticity of demand is instructive with regard to the standard practice in antitrust of defining markets by deciding whether particular products are sufficiently good substitutes—generally understood as having sufficiently high cross-elasticities of demand—to be included in the market. The expression makes clear that even a substitute with a very high cross-elasticity may have much less influence than that of a large group of other

²³ This point is elaborated in Kaplow (1982).

²⁴ Cross-elasticities need not be positive. For example, when the weighted summation equals zero, we have the familiar case of unit elasticity—that is, as price rises, expenditures on the product in question remain constant—and when the summation is negative, we have an elasticity less than one, often referred to as inelastic demand.

products, no one of which has a particularly high cross-elasticity. Moreover, products' shares of total revenues are not ordinarily considered in an explicit way, yet the formula indicates that a substitute with half the cross-elasticity of another can readily be more important, in particular, if its associated revenues are more than twice as high. More broadly, this representation of the relationship between overall elasticity and individual cross-elasticities reinforces the point that the effect of substitutes is a matter of degree and thus not well captured by the all-or-nothing approach involved in defining antitrust markets.

Some further comments concerning market share are in order, particularly in light of the fact that a persistently high market share is very frequently presented as compelling evidence that a firm has market power. No doubt this inference is often valid, specifically, if the market demand elasticity and rivals' supply elasticities are low in magnitude and the market conditions are reasonably stable. However, a firm with only a modest cost advantage may profitably maintain its high share by pricing low enough to capture most of the market. This occurs, for example, in the model of the dominant firm facing a competitive fringe if the fringe supply is very elastic at a price just above the firm's own marginal cost. Consider, for example, a trucking firm that provides 100% of the freight transportation on a particular route but would quickly be displaced by nearby rivals (whose costs are essentially the same but who suffer a slight disadvantage due to a lack of familiarity with the route's customers) if it were to raise its price even a few percent. Additionally, a firm may have a 100% share in market protected by a patent, but if there are sufficiently close substitutes, its market power is negligible. Conversely, even a firm with a low share of sales of a particular product may have quite a bit of technical market power if the magnitude of the market elasticity of demand and rivals' elasticity of supply for that product is very low. Gasoline refining and electricity generation are two examples of products for which this latter situation can arise. In sum, the right side of expression (2) indicates that market share is only one factor that determines the elasticity of demand facing a firm, so the magnitude of market share is a relevant component of market power but not a conclusive indicator.

c) Rivals' Supply Response: Barriers to Expansion, Mobility, and Entry

In examining the right side of expression (2) for the firm's elasticity of demand, the preceding section focused on the market elasticity of demand and market share. However, the elasticity of supply by rivals is also relevant, as indicated by the just-mentioned contrasting examples of trucking, on one hand, and gasoline refining and electricity generation, on the other hand. The concept of rivals' supply should be understood broadly, to include expanded output from existing plants, shifting capital from other regions or from the production of other products, introducing new brands or repositioning existing ones, and entry by firms in related businesses or by other firms. If market power is significant, it must be that the aggregate of these potential supply responses—often referred to as expansion, mobility, and entry—is sufficiently limited, at least over some pertinent time period. Gilbert (1989) provides an extended discussion of such barriers, Berry and Reiss (2007) survey empirical models of entry and market power, and Sutton (2007) discusses the relationship between market structure and market power.

In some cases, the elasticity of rivals' supply may be measured directly, by measuring output responses to previous changes in price by the firm in question, or by other firms in similar markets. Often, however, some extrapolation is required, such as in predicting whether a

hypothetical increase in price to unprecedented levels following a merger would generate a significant supply response. For internal expansion by existing rivals, the question would be whether there exist capacity constraints, steeply rising marginal costs, or limits on the inclination of consumers of differentiated products to switch allegiances. In the case of new entry, timing, possible legal restrictions (intellectual property, zoning, and other regulatory constraints), brand preferences, the importance of learning by doing, and the ability to recoup fixed costs, among other factors, will determine the extent of restraint imposed.

Particularly regarding the latter, it is common to inquire into the existence of so-called barriers to entry (sometimes taken as a shorthand for all forms of supply response by rivals). In some instances, such as when there are legal restrictions, the meaning of this concept is fairly clear. However, in many cases, it is difficult to make sense of the notion of entry barriers in a vacuum. For example, there is much debate about whether economies of scale should be viewed as a barrier to entry. If minimum efficient scale is large and incumbent producers have long-term exclusive dealing contracts with most distributors, entry may be rendered too costly, and existing firms might enjoy high price-cost margins (more than necessary to cover fixed costs). If instead there merely exist fixed costs and marginal costs are constant, in a free-entry equilibrium there will be positive price-cost margins yet no profits. The positive margins will not induce further entry because their level post-entry would be insufficient to recover fixed costs. As we have observed repeatedly, although market power would exist in the technical sense, the situation should not be viewed as problematic from an antitrust perspective.

Many structural features of markets have been identified as possible entry barriers: economies of scale, learning –by doing, reputation, access to capital, customer switching costs, lack of product compatibility, network effects, patent protection, and access to distribution channels. Because the implication of so-called entry barriers depends on the context—and because some degree of market power is sometimes unavoidable yet many are reluctant to state or imply its existence, such as by deeming something to be an entry barrier in a setting where antitrust intervention seems inappropriate—there is no real consensus on how the term “barriers to entry” should be defined or applied in practice.²⁵ We do not see clear benefits to formulating a canonical definition of the concept. It may be best simply to keep in mind the purpose of such inquiries into the existence of entry barriers: to assess rivals’ supply response as an aspect of an inquiry into the existence of market power, noting that market power is often relevant to antitrust liability but not sufficient to establish it. Beyond that, it may be more helpful to defer further analysis until considering specific practices in specific settings.

3. Conduct

In some situations, one may be able to infer the presence of market power from the challenged conduct itself. If we observe a firm engaging in a practice that could not be profitable unless it enhanced the firm’s market power to some certain degree, we may then infer that market power would indeed increase to that degree. For example, if a firm pays large amounts to retailers to agree not to deal with prospective entrants or spends large sums to maintain tariffs, we may infer

²⁵ See McAfee, Mialon and Williams (2004), Carlton (2004), and Schmalensee (2004) for recent discussions of how to apply the concept of entry barriers in antitrust analysis.

that these practices create or enhance that firm's market power.²⁶ If one accepts the premise that a firm's expertise in assessing its own market power is likely to be more reliable than that produced by a battle of experts before an agency or in litigation, then the firm's own conduct may be a sound basis for inferring the existence of market power.

Two caveats should be noted. First, the amount of market power that may be inferred will sometimes not be very great. A firm with billions of dollars in sales would happily spend millions lobbying for tariffs even if the resulting degree of market power were trivial. On the other hand, if a firm engages in a plan of below-cost pricing that sacrifices hundreds of millions in current profits, in the absence of other explanations one might well infer that it anticipates a substantial degree of market power, at least sufficient to recoup its investment.

Second, the reliability of the inference depends greatly on the lack of ambiguity regarding the character of the practice under consideration. If one is certain that the conduct would only be undertaken if it could enhance the firm's market power (to some requisite degree), then the inference is sound. However, often it will be contested whether the conduct in question was designed to and will have the effect of increasing market power rather than constituting a benign or even beneficial practice that increases welfare. For example, prices below cost may be profitable because they are predatory, or because they are introductory offers that will enhance future demand for an experience good, or because they stimulate the demand for other products sold by the firm at a healthy price-cost margin. If pro-competitive explanations are sufficiently plausible, no inference of market power is warranted, at least without further investigation.

Recognizing the possibility that the conduct at issue may be pro-competitive is especially important given the role that market power requirements often play in antitrust, namely, as a screening device. That is, we may require a plaintiff to prove the existence of market power because we do not want to subject a wide range of behavior to the costs of antitrust scrutiny and the possibility of erroneous liability. When the conduct that provides the basis for inferring market power is the very same conduct under scrutiny, and furthermore when the purpose and effect of such conduct is ambiguous, permitting an inference of market power from the conduct somewhat undermines the screening function of the market power threshold. This concern may be especially great when juries serve as the finders of fact.²⁷

²⁶ As we discuss in subsection V.D.2 in our analysis of exclusive dealing contracts with retailers, we would need to rule out pro-competitive justifications, such as those based on free riding. In the case of lobbying to erect tariff barriers, even if the conduct enhances market power, it would not violate U.S. antitrust laws because petitioning government, even to restrict competition, is exempt activity under the *Noerr-Pennington* doctrine.

²⁷ This concern may help to explain the Supreme Court's decision in *Spectrum Sports v. McQuillan*, 506 U.S. 447 (1993), where the Court held in an attempted monopolization case that the plaintiff had to meet the market power requirement independently of proving predatory conduct. Although the holding on its face seems illogical (if, as the plaintiff argued, it would have been irrational to have engaged in the conduct unless the requisite contribution to market power were present), the actual practice under consideration may well have appeared to the Court to be nonpredatory, so it wished to heighten the plaintiff's required proof before it would allow the case to be considered by the jury.

E. Market Power in Antitrust Law

As noted, in antitrust law the notion of market power is frequently used as a screen: a firm (or group of firms) must be shown to have some level of market power as a prerequisite to considering whether the conduct in question gives rise to antitrust liability. As a result, antitrust investigations and adjudications devote substantial attention to whether or not the requisite market power exists. In rhetoric and often in reality, this legal approach of viewing market power as something either present or absent—a dichotomous classification—is at odds with the technical economic notion of market power as a matter of degree. Because some degree of technical market power is ubiquitous, it is evident that the term “market power” as used in antitrust law has another meaning. Nevertheless, the law’s notion of market power is quite closely related to that of economists. A legal finding of market power constitutes not merely a declaration of the existence of technical market power, however trivial, but rather a conclusion that the degree of existing or increased market power exceeds some threshold, a benchmark that as we will see varies with the type of conduct under consideration and that in most instances is not clearly specified.

This feature of antitrust law’s use of a market power requirement is well illustrated by the law of monopolization. As will be elaborated in Section V.B, under U.S. antitrust law “[t]he offense of monopoly . . . has two elements: (1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident.”²⁸ The requirement of “monopoly power” is conclusory in that it merely signifies that degree of market power deemed minimally necessary and also sufficient to satisfy the first element of the offense of monopolization. It is understood that this level of market power is higher than that required in other areas of antitrust law. Notably, the market power requirement is highest in monopolization cases, somewhat lower in attempted monopolization cases, and lower still in horizontal merger cases, as will be discussed in subsections IV.D.2 and IV.B.2. However, these requirements typically are not stated quantitatively, making it difficult to know very precisely what is the threshold in any of these areas.

In principle, the fact that market power is a matter of degree should be recognized in designing antitrust rules. A monopolistic act that is unambiguously undesirable might be condemned even if the incremental impact on market power is modest, whereas for conduct that is ambiguous, with a high risk of false positives, it may be appropriate to contemplate condemnation only when the potential effect on market power is substantial. If one were minimizing a loss function in which there was uncertainty about the practices under scrutiny, and if the degree of harm conditional on the practices being detrimental was rising with the extent of market power, an optimal rule could be stated as entailing a market power requirement that was highly contextual.

For practical use by agencies and in adjudication, however, a more simplified formulation may economize on administrative costs, provide clearer guidance to parties, and reflect the limited expertise of the pertinent decision-makers. Nevertheless, some greater flexibility may be

²⁸ United States v. Grinnell Corp., 384 U.S. 563, 570-71 (1966).

warranted and is indeed increasingly reflected in antitrust doctrine. The early emergence of a *per se* rule against price-fixing, which dispenses with proof of market power, is one illustration. Another is the increasing use of intermediate levels of scrutiny under the rule of reason (see subsection III.E.2) and the implicit reliance on different market power thresholds by the antitrust agencies in reviewing horizontal mergers in different industries, despite the existence of official guidelines that purport to be uniform.

In addition to differences in the magnitude of market power thresholds and whether there is some flexibility regarding the requisite degree of market power, there is variation across contexts in whether the question posed concerns the extent level of market power or the amount by which the actions under scrutiny would increase it. In a monopolization case, the standard question is often whether a firm's past practices have improperly created or maintained monopoly power, so the inquiry is usually into whether significant market power already exists, as reflected in the previously quoted formulation. By contrast, in examining horizontal mergers, the focus is on whether the proposed acquisition would significantly increase market power.²⁹

We believe that this distinction is overstated and potentially misleading and that the correct inquiry should focus largely on contributions to market power. Even in the typical monopolization case, the relevant question is how much the past practices contributed to the existing situation. If the contribution is large (and if the practices are not otherwise justifiable), it seems that there should be a finding of liability even if the resulting total degree of power is not overwhelming. (In such a case, the initial level of market power presumably will have been rather low.) Likewise, even if the degree of existing market power is great, in cases in which the practices in question did not plausibly contribute significantly to that result, one should be cautious in condemning those practices, that is, they should be condemned only if they are unambiguously undesirable.

As an example, consider a firm selling a relatively homogeneous product, such as in the chemical industry, that enjoys a significant cost advantage over its rivals based on patented process technology. That firm might well enjoy a nontrivial degree of technical market power. Neither good sense nor existing law ordinarily condemns the discovery of a superior production process. Let us assume that the firm's technical market power was legally obtained and suppose further that the firm prices against a perfectly elastic rival supply at some trigger price that is below the firm's monopoly price. Antitrust issues could arise if this firm attempts to acquire its rivals or if the firm engages in conduct that drives its rivals out of business. In considering such cases, the degree of the firm's initial market power is of secondary importance (although if it were near zero, further inquiry would probably be pointless). Instead, the central question should be whether and to what extent the acquisition or exclusionary conduct will augment that firm's market power and thus harm consumers. For example, however great is the initial level of market power, the firm would gain no additional power by acquiring (or destroying) one of its rivals as long as numerous others that remain still have highly elastic supply at that same trigger price. However, the firm might well gain market power by acquiring or destroying a rival with uniquely low costs, thereby raising the price at which substantial competing supply would be

²⁹ See subsection IV.D.2, where we discuss the point that the extant level of market power is also important.

triggered. We return to the question of the relevance of extant market power versus challenged practices' contribution to power in subsection V.A.2 with regard to monopolization and exclusionary practices.

A further possible deviation between economic analysis and antitrust law with regard to market power concerns the benchmark against which the height of price-cost margins is assessed. The U.S. antitrust enforcement agencies in the Horizontal Merger Guidelines (1992) define market power as “the ability profitably to maintain prices above competitive levels for a significant period of time,” and the Supreme Court has similarly stated that “As an economic matter, market power exists whenever prices can be raised above the levels that would be charged in a competitive market.”³⁰ If one understands the competitive price to refer to the price that would be charged in a hypothetical, textbook, perfectly competitive market in which firms have constant marginal costs equal to the marginal cost of the firm in question at the prevailing equilibrium, then the legal and economic concepts are essentially the same. However, the hypothetical competitive scenario that underlies such statements is rather vague: the counterfactual is not explicit, and some specifications that may implicitly be contemplated may not yield sensible answers. For example, what is meant by the perfectly competitive price in a market with fixed costs?

Courts have struggled with these issues for many years. The Supreme Court has stated that “Monopoly power is the power to control prices or exclude competition.”³¹ This is not a meaningful screen, however, since any firm with technical market power has some ability to control prices. Conversely, in the European Union, the European Court of Justice has said that a “dominant position” corresponds to “a position of economic strength enjoyed by an undertaking which enables it to hinder the maintenance of effective competition on the relevant market by allowing it to behave to an appreciable extent independently of its competitors and customers and ultimately of consumers.”³² This test is not especially useful, either, since even a firm with great market power does not rationally behave independently of its competitors or customers. That is, there is some monopoly price, P_M , which—however high it may be—implies that a price of, say, $2P_M$ would be less profitable due to far greater consumer substitution away from the product at that higher price.

III. Collusion

We now turn to collusion, including price-fixing cartels and other arrangements that may have similar effects, such as the allocation of customers or territories to different suppliers.³³ For concreteness, we will ordinarily focus on price-fixing. There is an enormous literature on the

³⁰ Jefferson Parish Hospital District No. 2 v. Hyde, 466 U.S. 2, 27 n.46 (1984).

³¹ U.S. v. E.I. du Pont de Nemours & Co, 351 U.S. 377, 391 (1956).

³² Case 322/81, Michelin v. Commission [1983] ECR 3461 § 30.

³³ We do not explicitly address the full range of “horizontal agreements,” which includes group boycotts as well as arrangements among buyers, notably, to suppress the prices of inputs, the latter of which are subject to similar analysis as that presented here, although they have received less antitrust scrutiny.

economics of collusion that we do not attempt to review systematically. Existing surveys include Shapiro (1989), Jacquemin and Slade (1989), Motta (2004, ch. 4), and of particular note Whinston (2006, ch. 2). For in-depth discussion of some especially interesting price-fixing cases, see Borenstein (2004), Connor (2004), Elzinga and Mills (2004), Motta (2004, pp. 211-219), and Porter and Zona (2004).

The focus here, as in the rest of this survey, is on the intersection of economics and the law. We begin by noting the core elements from each field and posing questions about their relationship. Next, we explore the economics of collusion, focusing on the necessary elements for successful collusion, lessons from game-theoretic models of oligopoly, and the various factors that bear on the likelihood of successful collusion. Finally, we examine legal prohibitions in light of the basic teachings of economics.

A. Economic and Legal Approaches: An Introduction

1. Economic Approach

For as long as there has been commercial competition, rivals have been tempted to short-circuit it because self-interest favors their own profits at the expense of customers' interest in lower prices and the overall social interest in allocative efficiency. No less a champion of the free-market system than Adam Smith ([1776] 1970, bk 1, ch. X) considered collusion an ever-present danger. "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." If one thinks in terms of a homogeneous product, firms seek to establish and maintain the monopoly price, which exceeds the price that would prevail in the absence of the agreement. With differentiated products or price discrimination, although there is no single monopoly price, the same idea applies: firms seek to elevate prices and thus raise their collective profits at the expense of consumers. In so doing, the firms typically increase the gap between price(s) and marginal cost(s) and thus raise deadweight loss and lower total welfare, defined as the sum of supplier profits and consumer surplus. Thus, collusion is unwelcome, whether one is seeking to maximize overall efficiency or consumer welfare.

Colluding firms use a variety of methods to achieve the basic goal of raising prices. In some cases, firms agree to minimum prices. In others, they agree to limit their production levels, since output restrictions translate into elevated prices. Alternatively, firms can allocate customers or territories among themselves, with each firm agreeing not to compete for customers, or in territories, assigned to others. These customer and territorial allocation schemes effectively grant each firm a monopoly over some portion of the overall market, so they lead to higher prices and reduced output, even though these schemes do not directly specify price or output.

Economists studying collusion, and more generally oligopoly, tend to inquire into the factors that determine the market equilibrium outcome in an industry. Economists typically focus on whether the outcome is relatively competitive, with prices close to marginal cost, or at least some measure of average cost, or relatively collusive, with prices close to the level that a monopolist would pick to maximize industry profits. This approach is consistent with economists' traditional emphasis on market outcomes and their implications for the allocation of resources.

This approach focuses on description or prediction, not on policy prescriptions regarding how the government should mitigate the costs of collusion. There is a general consensus that clearly identifiable attempts to engage in collusive behavior should be prohibited, so explicit cartel agreements should not be legally enforceable and private attempts to agree upon and enforce supra-competitive prices should be punished. It is widely recognized, however, that it is not always possible to determine whether collusion is occurring or, even when it is, which specific practices should be proscribed. One approach in such settings would be price regulation, which is often undertaken in the case of natural monopolies but is generally thought to be inferior to decentralized competition when such is feasible. In the past, there have been recommendations to deconcentrate certain industries in order to achieve more competitive outcomes.³⁴ Such proposals have not been implemented, except in some cases of monopolization, and have not of late been actively considered in the United States. Another structural approach is more prevalent: enjoining horizontal mergers that make collusive outcomes more likely, a topic we explore in Section IV. Finally, for cases in which collusion can be identified but the specific practices enabling it cannot, Posner (1969, 2001) interestingly proposes the imposition of monetary penalties on oligopolists if the market equilibrium outcome is collusive.³⁵ The idea is that, just as Pigouvian taxes induce firms to refrain from inefficient behavior the details of which might be difficult for a regulator to observe or proscribe, so too would appropriate fines or damage awards in private litigation lead firms to abstain from collusive behavior. This approach assumes, importantly, that it is possible to measure the extent to which prices exceed non-collusive levels, which poses both conceptual and practical challenges of a sort that are encountered in imposing sanctions, the magnitude of which depends on the extent of collusive overcharges. However, Posner's approach has not been embraced by the courts.

2. Legal Approach

The legal approach to collusion, at least on its face, differs from the economic approach. As just described, the economic approach begins with a diagnosis of the problem, then tries to ascertain whether and when collusion occurs, and finally assesses the efficacy of competing remedies. Although one would like to believe that the legal approach is at some level grounded in such analysis, on the surface it appears to focus instead on particular behavioral elements. As will be seen in the course of our analysis in Section III, the extent to which the legal approach can ultimately be rationalized on economic grounds depends on how legal tests are interpreted.

³⁴ Legislation was introduced repeatedly in the early 1970s that would have authorized the dissolution of firms in concentrated industries that had not engaged in substantial price competition over three consecutive years; see, for example, S.1167, March 12, 1972. This legislation was based on the White House Task Force Report on Antitrust Policy (1968), commonly known as the Neal Report.

³⁵ On its face, present practice appears to differ significantly from Posner's proposal. Although the United States and most other competition enforcement regimes do provide for fines or private damage remedies in cases of price-fixing, to trigger such payment obligations, the government or private parties typically need to show that in some sense there is an "agreement." Furthermore, satisfaction of this requirement is generally understood to entail more than demonstrating that the observed outcome involves a "collusive price," although as we shall discuss, just how much more must be shown and what constitutes an adequate demonstration is unclear. See subsection 2 and section D.

In the United States, the European Union, and many other jurisdictions, the structure of legal prohibitions revolves around the distinction between unilateral and group behavior. Unilateral behavior is circumscribed to a limited degree by anti-monopolization law (see Section V) and various other provisions but is not subject to a regime of price regulation or other forms of internal micro-management of firms' dealings.³⁶ Thus, firms are purportedly free to set prices and other conditions of trade.³⁷ This freedom, however, is restricted to unilateral behavior. Independent firms are expected to compete, conferring the benefits of competition on consumers and on society as a whole.

The central legal question with which we will be concerned—and will elaborate in section D—is how courts or other regulators are to determine when supposedly competing firms are instead conspiring. Legal prohibitions are typically triggered by certain types of conduct rather than by outcomes themselves. For concreteness, we will discuss the prohibition in U.S. antitrust law, Sherman Act §1, which makes illegal “[e]very contract, combination . . . , or conspiracy, in restraint of trade.” In practice, the standard term of art is “agreement,” even though that term does not appear in the statute.³⁸ Thus, the legal question is whether firms' pricing is a result of an agreement. If not, there is no violation. If so, there is a violation, and penalties in the United States include having to pay treble damages to injured customers, being subject to injunctions on prohibited behavior, and criminal penalties, under which firm executives convicted of price-fixing serve prison terms and firms pay fines.

What, then, is an agreement? And how does this concept relate to the economic analysis of collusive behavior? These questions will occupy much of the remainder of Section III of our survey. To provide some guidance in the interim, a few preliminary observations are offered. First, there are clear cases. At one extreme, if competitors meet in the proverbial smoke-filled room, negotiate a detailed cartel arrangement, sign it, and implement it—and, importantly, this all can be proved in a legal proceeding—an agreement and hence a legal violation will undoubtedly be found to exist. At the other extreme, no agreement would presumably exist and no violation would be found due to the mere fact that competitors' prices are equal—as one expects with homogeneous products and perfect competition, for example—or that they sometimes move together—as tends to occur when there are shocks to input prices (think of retail gasoline stations changing sale prices when prices from refineries change).

The difficult cases fall at various points in between, in terms of what actually transpired and what can be proved before a tribunal. Consider a simple example. Suppose that just two firms, A and B, supply a particular product. Let the monopoly price of that product be \$100 and the fully competitive price (that is, the price at which the industry marginal cost curve crosses the

³⁶ It is true, however, that remedies in monopolization cases and some others can entail what is tantamount to fairly detailed regulation.

³⁷ There are important qualifications, notably with regard to prescriptions on predatory pricing (see Section V.C), but the focus in this Part is on prices that are too high and thus directly harm customers rather than on prices that are too low and thus directly harm competitors.

³⁸ Interestingly, this is the language of Article 81 of the competition law in the European Union. Like in the United States, the concept in the European Union embraces more than formal contracts yet it is uncertain just how much more.

demand curve) be only \$40. Suppose further that the actual industry price persists at \$100, with sales split evenly between the two firms. This is clearly a collusive outcome, but have the firms entered into an agreement in restraint of trade?

As noted, if a written agreement, negotiated and signed in a smoke-filled room, is produced as evidence, a violation will be found. Suppose that no such agreement is directly in evidence. One possibility is that such an agreement nevertheless exists, and if a tribunal can be convinced of this by circumstantial evidence, a case will have been made. But what sort of evidence would be necessary to make this inference? The answer depends importantly on the competing hypotheses—and on which alternative explanations are likewise deemed to involve agreements and hence would also constitute violations. What interactions short of a meeting in a smoke-filled room that results in a written document will suffice? Is a face-to-face meeting required or would a conference call or an e-mail exchange be enough? What about other forms of communication, such as statements relayed through third parties or in various codes? Or nonverbal communication (hand signals, winks and nods, posting signs with proposed prices, and so forth)? Must there be a written document? Presumably not. Must there be a formal agreement tantamount to a legally enforceable contract? Well, since a contract would not be legally enforceable in any event, presumably this too is not required.

In sum, we can be certain that agreements may be deemed to exist when something well short of the formal meeting and written cartel document exists. But it is not clear how much less will give rise to liability, or, put in the affirmative, just what is (are) the core underlying element(s) of an “agreement.” For now, we will leave this question, returning to it in section D, after we have surveyed some key aspects of the pertinent economic theory, which one might hope would illuminate the legal inquiry.

B. Oligopoly Theory

1. Elements of Successful Collusion

Economists have long recognized that there exist certain prerequisites to successful collusion. The classic modern reference is Stigler (1964). Green and Porter (1984) embed these issues in a supergame context. The key elements are (1) *reaching consensus*: some understanding must be reached among the otherwise-competing firms regarding what conduct is permitted under the terms of the collusive agreement, such as the prices that the firms will charge; (2) *detection*: some reliable means must exist by which departures from the agreement can be detected; and (3) *punishment*: some credible mechanism must be established by which such departures are punished if and when they are detected. Specifically, the prospect of detection and punishment must be sufficient to deter individual firms’ proclivity to cheat on the agreement, typically by cutting prices in the short-term, hoping to reap greater profits through a higher market share, at the expense of the other firms, before they can respond. Related to the need to reach an agreement is the problem of (4) *inclusion*: a means of inducing participation by a sufficiently large number of incumbent suppliers so that competition from non-participants does not undermine the profitability of the collusive agreement. Lastly and relatedly, the incumbent firms must be protected by (5) *entry barriers*: there must not be so much competition from quickly arriving new entrants so as to undermine the effectiveness of collusion.

Some economists consider these requirements to be so daunting that cartels are either unable to form or quick to collapse, even in the absence of antitrust laws designed to stop collusion. For example, when OPEC first arose, some confidently predicted its immediate demise. However, the experience with OPEC and empirical evidence on price-fixing more broadly does not support this optimistic view. For example, in the past decade, the Antitrust Division of the U.S. Department of Justice has broken up many large, international cartels that had operated for years (despite the fact that they were illegal under the antitrust laws) and successfully reaped hundreds of millions if not billions of dollars in profits.³⁹ For recent extensive surveys of the evidence of collusive activity, see Connor (2007), Harrington (2006b), and Levenstein and Suslow (2006). Whinston (2006, pp. 26-38) offers a more selective discussion of the empirical evidence regarding the effects of price-fixing conspiracies.

2. Repeated Oligopoly Games and the Folk Theorem

The basic theoretical framework used to evaluate the presence, absence, or efficacy of collusion is that of dynamic or repeated oligopoly, that is, situations in which an identifiable group of suppliers offering substitute products interact over time.⁴⁰ This framework includes infinitely repeated oligopoly games, so-called supergames. Cartel theory requires dynamic analysis because the central elements of detection and punishment inherently take place over time.

One of the central findings in the theory of oligopoly supergames is that there are many—indeed, infinitely many—noncooperative equilibrium outcomes, including outcomes that maximize the joint profits of the oligopolists, even if one restricts attention to subgame-perfect Nash equilibria.⁴¹ To give a flavor for why there are so many equilibria in supergames, consider a game in which each of N firms, selling the same homogeneous good and incurring a constant cost per unit of C , sets its price every period. Suppose that the stage game is a classic Bertrand pricing game, so the firm with the lowest price in a given period serves the entire market that period, and if multiple firms charge the same lowest price output is shared equally among them. (One can also think of the firms bidding each period to serve the single large customer who buys that period.) For the moment, suppose also that, as soon as a given period ends, each firm immediately observes all the prices charged by the other firms during that period.

³⁹ See Litan and Shapiro (2002) for a discussion of U.S. cartel enforcement activities during the 1990s. The Antitrust Division of the Department of Justice regularly announces enforcement actions against cartels. For example, as of September 2006, some \$731 million of fines had resulted from the Division's investigation into price fixing of DRAM (Dynamic Random Access Memory). See "Samsung Executive Agrees to Plead Guilty, Serve Jail Time, for Participating in DRAM Price-Fixing Conspiracy," http://www.usdoj.gov/atr/public/press_releases/2006/218462.htm.

⁴⁰ There is a very large literature on oligopoly theory and supergames. Since our focus is on antitrust, we will draw on this literature but not review it in any systematic way. See Shapiro (1989) for a survey of oligopoly theory. Ivaldi, et. al. (2003a), Motta (2004, ch. 4), and Whinston (2006, ch. 2) provide recent discussions of the application of oligopoly theory to collusion cases in antitrust.

⁴¹ Friedman (1971) showed that the full cartel outcome can be supported in a repeated oligopoly if the players are sufficiently patient.

In this simple repeated Bertrand game, the competitive outcome, by which we mean the Nash equilibrium in the one-shot (stage) game, involves each firm setting its price equal to cost, C , each period. This would be the only noncooperative equilibrium outcome if this pricing game were played only a single time, or indeed any finite number of times (a familiar consequence of backward induction). However, if the game continues indefinitely, it is easy to construct a subgame-perfect equilibrium in which the price each period is $P > C$, for many different values of P ranging from the competitive price all the way up to the monopoly price (or, in fact, higher). The trick is to postulate that, should any firm ever charge below P , all the other firms will set a price of C in all subsequent periods. This punishment strategy thus entails reversion to the one-shot Nash equilibrium. Because this behavior supports an equilibrium price of P every period, the equilibrium profit of each firm is $\pi(P)/N$ each period, where $\pi(P)$ denotes the profits earned by a firm setting price P and serving the entire market. Adding up over all periods and discounting using the one-period discount factor δ , the equilibrium profits of each firm are $\pi(P)/N(1-\delta)$. In contrast, in any given period a single firm could defect by slightly discounting its price, thereby capturing the entire market that period. This would, by hypothesis, condemn all of the firms—importantly, including itself—to the competitive outcome for all future time. The payoff to this defecting firm is $\pi(P)$ in the current period and zero in all subsequent periods. Defecting is therefore unattractive if and only if $\pi(P)/N(1-\delta) > \pi(P)$, that is, if and only if $\delta > 1 - 1/N$. If the periods are sufficiently short, that is, if price cuts are detected rapidly enough, then δ is close to unity and this condition is met, even when N is large.

The frequency of sales, and hence the speed of detection, is implicitly built into this simple model. As expected, the faster that rivals can learn of one firm's defection and respond by reducing their own prices, the easier it is to sustain collusion. Formally, the model assumes that price cuts are observed after one period, but the length of time comprising one period has not been specified. If one period takes time T , and if the interest rate per unit time is r , then the discount factor is $\delta = e^{-rT}$. Therefore, a longer detection lag corresponds to a lower discount factor, making it less likely that collusion will be sustainable, *ceteris paribus*.

This highly simplified example illustrates that there exists a subgame-perfect equilibrium in which all the firms charge the same price, P , repeatedly, for any $P > C$, as long as detection is sufficiently rapid. For example, if the length of a period is one month, and we imagine that the interest rate is around 1% per month, then the discount factor δ is roughly 0.99, so the monopoly price can be sustained as a perfect equilibrium in the repeated oligopoly game as long as the number of firms is less than one hundred.

While the specific calculations here depend upon the particular oligopoly game being played each period, the basic idea—that there exist perfect equilibria with high (well above competitive) prices even with many firms for plausible discount factors—is by no means specific to this example. Consider, for example, Cournot oligopoly with a homogeneous product. On one hand, a firm that defects cannot gain nearly as much as in the Bertrand game. (The ratio of profits from defecting to the per-period profits from following the equilibrium strategy depends upon the shape of the demand curve.) On the other hand, reverting to the static Cournot equilibrium is not as severe a punishment as reverting to the static Bertrand equilibrium. Shapiro (1989) shows that, with $\delta = 0.99$, the monopoly price can be sustained in a repeated Cournot oligopoly with constant marginal cost and linear demand as long as there are no more than four hundred firms.

Furthermore—one might say “worse yet”—this example is hardly anomalous. To the contrary, the “folk theorem” for repeated games tells us that, quite generally, there exists a plethora of equilibria in repeated games, including equilibria that correspond to full cooperation (the profit maximizing cartel outcome), as long as the players are sufficiently patient, that is, if the discount factor is sufficiently close to unity. Fudenberg and Maskin (1986) provide general conditions under which any feasible and individually rational payoff vector in the stage game can be achieved as a perfect equilibrium in the repeated game if players are sufficiently patient.⁴² Fudenberg, Levine, and Maskin (1994) extend this result to games in which the players do not observe each others’ actions but only a public outcome (such as price) that signals those actions.

These “folk theorems” pose two related and fundamental challenges for the analysis of oligopoly. First, these strong results do not comport with the observation that rivals often compete rather than cooperate and in particular with the prevailing view that collusion is difficult if the number of firms is moderately large—a view that Farrell (2000) calls the “structural consensus.” Clearly, some important things are missing in models of repeated oligopoly that predict collusive outcomes for a very wide range of market structures and industry conditions. Put simply, supergame theory, at least with sufficiently patient players and without further modifications, is of limited use because it proves far too much. The economics literature has responded to this criticism by probing the assumptions underlying the folk theorem and by exploring equilibrium outcomes when the relevant discount factor is not very close to unity, that is, when the players are not “sufficiently patient.” In particular, a large literature, some of which we examine in section C, explores the conditions that make it more or less difficult to support collusive outcomes in repeated oligopoly.

Second, and of particular relevance to antitrust law, there is no explicit role for communications in the basic models of repeated games, so these models do not help us understand the impact of meetings and other communications among oligopolists. Some research, however, does explore aspects of this limitation, as we consider next.

3. Role of Communications

One of the specific shortcomings of standard models of oligopolistic supergames is that they do not help us understand how the firms initially determine which of the plethora of equilibria to play. One interpretation of these games is that the firms engage in extensive communications and perhaps negotiations *before* the game begins, in order to agree upon the equilibrium that they will play. Under this interpretation, the equilibria in oligopolistic supergames represent self-enforcing outcomes that can arise once an agreement is reached. The alternative explanation of the observed conduct is that the firms somehow find their way to a relatively collusive outcome

⁴² Abreu, Pearce, and Stacchetti (1986, 1990) consider punishments stronger than simply reverting to the one-shot Nash equilibrium. In the simple repeated Bertrand game, no punishment can be stronger than reversion to the one-shot Bertrand equilibrium, since it involves zero profits, which by definition is the level of profits a firm can obtain by exiting. For the repeated Cournot game, however, stronger punishments are possible by targeting the defecting firm and punishing again a firm that refuses to go along properly with its own punishment. When available, these stronger punishments make it possible to sustain the monopoly price even if the firms are somewhat less patient.

without engaging in any communications, other than through their actions in the market, such as their setting of prices.

Ambiguity about the role of communications is inherent in the standard solution concept of Nash equilibrium (and thus perfect equilibrium, a refinement of Nash equilibrium). When the strategies are highly complex, and especially when there are multiple equilibria, the perfect equilibrium (or Nash equilibrium) methodology does not explain how the firms were able to coordinate to select an equilibrium. Yet common sense indicates that communications can play a role in such coordination, and complex strategies supporting a collusive outcome would seem to constitute evidence in favor of the hypothesis that the firms in fact met and reached some sort of agreement at some point in time.

The literature on “cheap talk” asks whether communications affect the equilibrium outcome in a game-theoretic setting. Farrell and Rabin (1996) provide an excellent overview of this literature.⁴³ In general, cheap talk, that is, communications that do not directly affect payoffs, can affect equilibrium outcomes.⁴⁴ Farrell and Rabin give the example of one Cournot duopolist telling another: “You cut your output and I’ll cut mine.” While this might be a trick—the speaker gains if the listener cuts output, whether or not the speaker does so—this also might be an effective way of initiating output reductions that sustain collusion in repeated play. In the end, Farrell and Rabin conclude that cheap talk about intended play can make a big difference when the players’ interests are well aligned, but the gains available from coordination can easily be lost due to dispute and bargaining problems.

A different strand of the literature studies the role of communications to convey private information. Athey and Bagwell (2001, 2004) permit the firms to communicate about private cost information in a repeated pricing game. Compte (1998) and Kandori and Matsushima (1998) study communications when firms observe private but imperfect signals about past play. Athey, Bagwell, and Sanchirico (2004) study a model with private cost shocks and publicly observed prices.

Facing the rather ambiguous theoretical results in the “cheap talk” literature reported by Farrell and Rabin (1996), economists have conducted experiments to learn how communications affect the strategies adopted by players. This literature is surveyed by Crawford (1998). As an early example, Cooper et. al. (1989) find that communications in the “battle of the sexes game,” where the players can gain from cooperation but do not agree about which outcome is best, greatly increase the chance that the players will successfully coordinate. Kuhn (2001) discusses the antitrust implications of this experimental literature, emphasizing the role of communications in achieving coordination by reducing uncertainty about what other firms will do.

⁴³ Much earlier, Schelling (1960) recognized the importance of communications and discussed the role of “focal points” in coordinating outcomes in strategic situations.

⁴⁴ Communications also may have no effect. There always exists a “babbling” equilibrium in which players ignore the statements made by others. As emphasized by Farrell and Rabin (1996), however, many of these equilibria are implausible.

The critical role of communications in sustaining collusion is revealed in the fascinating study by Genesove and Mullin (2001) of the Sugar Institute, a trade association that operated from 1927 to 1936. They examine in detail how sugar refiners established a set of rules to facilitate collusion. The Sugar Institute experience shows how weekly meetings among sugar refiners were used to establish and interpret rules that enforced business practices making price-cutting more transparent. In contrast to the theories described above, which involve no cheating in equilibrium, cheating did occur, but retaliation was carefully limited. This example illustrates a number of functions served by regular communications, functions that could not be served simply by initial communications. Related lessons can be found in Harrington (2006b), who reports on some twenty European Commission cartel cases from 2000 to 2004.

C. Industry Conditions Bearing on the Likelihood of Collusive Outcomes

As noted, to get beyond the folk theorem, the theoretical literature on oligopoly has extensively explored the conditions under which the joint profit-maximizing outcome can be achieved as a perfect equilibrium in a repeated oligopoly game when the discount factor is *not* close to unity. One way this question is posed is to ask how various factors affect the critical discount factor, δ^* , such that the fully collusive outcome can be supported as a perfect equilibrium for $\delta \geq \delta^*$.⁴⁵ While a survey of the enormous literature on oligopoly theory is well beyond the scope of this chapter, we mention here selected results that are especially relevant to antitrust. We also examine a number of other factors, not strictly part of the standard oligopoly supergame framework, that bear on the feasibility of collusion.

In the first three subsections, we relax several extreme and unrealistic assumptions made in the simple model used above. First, the simple model assumed that a defecting firm could capture the entire market with even a slight price cut. Second, it assumed that even a tiny price cut would surely be observed by the rivals. Third, it assumed that even the slightest defection would be punished severely, with all firms pricing at marginal cost in perpetuity, leading to zero profits for all firms. We then consider a variety of other factors that make it more or less difficult for collusive outcomes to be sustained as perfect equilibria in repeated oligopoly.

1. Limited Growth for Defecting Firm

There are many reasons why a firm that defects from collusive prices may not be able to capture the entire market, including upward sloping marginal cost (in the limit, capacity constraints), customer loyalty, customer switching costs, and product differentiation. Clearly, if the gains from defection are limited, collusion will be easier to sustain, *ceteris paribus*.

To illustrate this idea in the simple model introduced above, suppose that capacity constraints only permit a single firm to grow its sales by a factor $1 + g$ in a single period.⁴⁶ In the simple model above, $1 + g = N$, but if a firm can only, say, double in size in one period, then $g = 1$. For

⁴⁵ Ivaldi et. al. (2003a) take this approach in their accessible and informative overview paper.

⁴⁶ A fully specified model would relate this growth limit to underlying economic variables, such as capacity or the degree of product differentiation, and to magnitude of the defecting firm's price cut.

now, we retain the assumption that all price cuts are detected by rivals and punished strongly, leading to zero future profits for the defecting firm. With these assumptions, optimal defection involves a tiny price cut and yields profits of $\pi(P)(1+g)/N$. As a result, charging the collusive price is optimal so long as $\delta > \delta^* = \frac{g}{1+g}$. The smaller is g , the smaller is δ^* , indicating that collusion is easier to sustain. For example, if $g = 1$, then $\delta^* = 0.5$, compared with $\delta^* = 0.9$ in the case where $N = 10$ and a defecting firm could capture the entire market. However, the $\delta^* = 0.5$ calculation makes the assumption that a tiny price cut by one of ten firms, doubling its market share from 10% to 20%, would surely be observable to all of the other firms.

2. Imperfect Detection

As emphasized by Stigler (1964), Green and Porter (1984), and much of the subsequent literature, when one firm cuts its price, its rivals may not be able to observe the price cut. Imperfect detection unquestionably makes it more difficult to sustain collusive outcomes. In fact, collusion cases in antitrust law frequently revolve around whether the firms have sufficient ability accurately to observe price cuts so as to enforce a collusive outcome. Additionally, many of the factors considered in subsequent subsections take on importance because they affect the ability of the firms to detect and punish those who defect from the collusive arrangement.

To illustrate the fundamental importance of detection, consider how the calculus of defection changes if there is an exogenous probability, θ , that the price cut is observed by rivals.⁴⁷

Retaining our assumption that a single firm can only grow its sales by a factor $1+g$ in a single period, the payoff from cutting price for a single period is $\frac{\pi(P)}{N}(1+g) + (1-\theta)\frac{\pi(P)}{N}\frac{\delta}{1-\delta}$.⁴⁸

Collusion is sustainable if and only if this payoff is less than that from maintaining the collusive price, $\frac{\pi(P)}{N}\frac{1}{1-\delta}$. Simplifying, collusion is sustainable if and only if $\delta > \delta^* = \frac{g}{\theta+g}$. This

expression captures a basic tradeoff: collusion cannot be sustained if detection is very unlikely (low values of θ), especially if a firm can grow rapidly before detection would take place (large values of g). With $g = 1$ and $\theta = 0.25$, $\delta^* = 0.8$, far higher than $\delta^* = 0.5$ when $\theta = 1$.

Stigler (1964) emphasizes the role of price transparency and secret price cutting. Spence (1978) argues that uncertainty about demand conditions makes it more difficult for suppliers to distinguish shifts in demand from defections by their rivals, and thus makes collusive outcomes more difficult to sustain. Green and Porter (1984) derive trigger strategies when prices are only observed with noise, in which case there is a tradeoff: entering the punishment phase more readily provides a stronger deterrent but can lead to price wars even when no firm has defected. Harrington and Skrzypacz (2007) study a model in which sustaining collusion in the presence of

⁴⁷ Immediately below, we explain how the analysis changes if the probability of detection is a function of the price charged by the price-cutting firm.

⁴⁸ This expression assumes, as above, that the firm earns zero profits in the future if its price cut is detected, but also that if the price cut is not detected promptly, it is never observed.

demand uncertainty requires asymmetric punishments if the firms observe each other's outputs but not their prices.

In the defection calculations presented so far, there was no reason for a defecting firm to cut its price by more than a small amount since deeper price cuts were not needed for that firm to sell either to the entire market (in our initial model) or to as many customers as the firm can serve given its capacity (in the modified model). The defection calculations are more complex if the firm-specific demand curve is such that a defecting firm's profits are decreasing at the collusive price, in which case the firm's immediate profits would be higher with a discrete rather than incremental price cut. This variation alone could easily be accommodated using standard pricing theory: the defecting firm would maximize its immediate profits by pricing at the point where marginal revenue and marginal cost are equal.

The analysis becomes more complex, and more interesting, however, if we combine this idea with imperfect detection. More specifically, suppose that deeper price cuts are more likely to be detected by rivals. Then optimal defection involves a tradeoff: lower prices lead to higher profits in the immediate term but a higher probability of detection. The lesson is that the sustainability of collusion may depend in a complex way on the interaction between the ability of a defecting firm to gain customers in the short run and the ability of that firm's rivals to detect that the firm has departed from collusive behavior.

3. Credibility of Punishment

The simple model effectively assumed that it was credible for the oligopolists, as a group, to punish any defection by reverting in perpetuity to the "competitive" outcome, defined as the one-shot equilibrium. In the case of a pricing game with homogeneous products, this implied zero profits for all of the firms. Clearly, this is an extreme assumption, one that seems to dismiss the temptation of the firms to relent on their punishments and "try again" to achieve a more profitable collusive outcome.

To explore the role of the punishment strategies, consider first the importance of the magnitude of punishment. Suppose that, after one firm defects, instead of prices forever after being set at cost, the other firms respond by merely matching the initial price cut. With this assumption, a firm that cuts price to $P_{cut} < P$ earns total profits of $\pi(P_{cut}) + \frac{\delta}{(1-\delta)} \frac{\pi(P_{cut})}{N}$. This expression must exceed the profits of $\frac{\pi(P)}{N} + \frac{\delta}{(1-\delta)} \frac{\pi(P)}{N}$ from maintaining the collusive price. But profits from defection will indeed exceed the profits from continued collusion for values of P_{cut} near P since cutting price increases the first term by a factor of N and has an effect on the second term that vanishes as $P_{cut} \rightarrow P$. After all, the full benefit of defection—capturing the entire monopoly profit—is obtained with a small price cut in the period of defection, and the future punishment is negligible since the price only declines very slightly from the initially collusive price. This analysis illustrates that threats to merely match price cuts are insufficient to maintain collusion.

Second, consider whether the punishment of cutting price to cost in all future periods is credible. One might think that it is because the strategy consists of playing the stage-game Nash

equilibrium over and over again, and by construction no single firm would find it optimal to depart from such conduct, given the behavior of the other firms. However, there is something fishy about strategies that call for smooth initial coordination followed by a perpetual price war if any one firm departs from the agreed-upon price, however brief the period of time. In particular, would the firms not be tempted to relent on the price war at some point and return to cooperation? The tension arises because the logic of perfect equilibrium does not consider collective departures from the punishment regime. But ruling out collective departures is hard to defend in a theory that postulates from the outset that the firms can find a way to coordinate to select an equilibrium that is mutually beneficial.

One simple but *ad hoc* way of dealing with this point is to limit the duration of punishments that are allowed after one firm defects by cutting its price below the initially specified level. In the simple supergame above, suppose then that punishments are limited to K periods. With this limitation, the payoff to a single firm from defecting is equal to $\pi(P) + \delta^{K+1}\pi(P) / N(1-\delta)$, where the first term represents the profits during the period when the firm defects and captures the entire market and the second term measures the profits this firm earns once the collusive outcome is restored after K periods of punishment (during which no profits are earned). For collusion to be sustainable, this expression must be no larger than the profits from indefinitely charging the collusive price, P , which as before equal $\pi(P) / N(1-\delta)$. Continuing to use $\delta = 0.99$, for moderate values of K , this condition implies that collusion lasting K periods is sustainable with up to roughly $N \approx K$ firms.⁴⁹ So, if we think of one period as corresponding to one month and if we believe that the firms can credibly enter into a price war for one year following a defection, collusion is sustainable so long as there are no more than twelve firms.

A much deeper, but far more complex, way of addressing the credibility of punishments can be found in Bernheim and Ray (1989) and Farrell and Maskin (1989). These papers have proposed refinements of the perfect-equilibrium solution concept that rule out continuation play, including punishment strategies, that is not collectively credible. At the very least, a perfect equilibrium is rejected if the continuation play in any subgame is Pareto dominated by the continuation play in any other subgame. Using the terminology from Farrell and Maskin, who study two-person games, an equilibrium that does *not* contain one subgame that is Pareto dominated by another subgame is called “weakly renegotiation proof” (WRP), a refinement of the subgame-perfect equilibrium concept. In the example above, the continuation game after one firm defected involved all firms pricing at cost forever. In this subgame, all firms earn a continuation payoff of zero. This subgame is thus Pareto dominated by the subgame consisting of the game itself, in which each firm earns a payoff of $\pi(P) / N(1-\delta)$. Therefore, the perfect equilibrium used above to support the monopoly price is not weakly renegotiation proof, even though it is

⁴⁹ Defection is unattractive so long as $\pi(P) + \delta^{K+1}\pi(P) / N(1-\delta) < \pi(P) / N(1-\delta)$ which is equivalent to $N(1-\delta) + \delta^{K+1} < 1$. For values of δ near one, $\delta^K \approx 1 - K(1-\delta)$, so this expression is approximately $N(1-\delta) + 1 - (K+1)(1-\delta) < 1$, which can be written as $N-1 < K$. Intuitively, defecting gives the firm an extra $N-1$ times its profits right away, which is balanced against the loss of those profits for K periods. For $\delta = 0.99$ and moderate values of K , a good approximation for the maximum number of firms is $N \approx K$.

subgame perfect. All of the firms would prefer to resume cooperating rather than carry out the punishment.

Farrell and Maskin (1989) show that the WRP condition, when applied to repeated-duopoly Bertrand and Cournot games, rules out highly asymmetric equilibria, even if the discount factor is close to unity. However, the WRP condition does not rule out many equilibria involving the monopoly price, including the symmetric equilibrium in which the two firms split the monopoly profits. If asymmetric punishments are specified that favor one firm, that firm will block renegotiations, which WRP required to be a Pareto improvement. Therefore, the WRP concept alone does not successfully resolve the paradox associated with the folk theorem as applied to oligopolies. Farrell (2000) suggests a further refinement, which he calls quasi-symmetric WRP, motivated by the notion that all innocent firms will be treated symmetrically. This concept requires that all innocent firms prefer to carry through with the punishment of a defecting firm rather than revert back to the original equilibrium strategies. Farrell shows that monopoly prices cannot be supported, regardless of the discount factor $\delta < 1$, for moderate numbers of firms in Bertrand or Cournot oligopoly, if this condition must be satisfied. This approach is promising, but relies on the assumption that the firms would find it difficult to establish punishments that treat innocent firms asymmetrically. Further work is required before these promising ideas can be put to practical use to help identify industry conditions under which collusion is most likely to be effective while accounting for the collective credibility of responses to defections.⁵⁰

4. Market Structure

We now consider a series of factors relating to market structure that affect the incentive and ability of oligopolistic suppliers to sustain a collusive outcome in repeated play.

a) Market Concentration

Collusive outcomes are less likely to occur in industries with more firms because greater numbers make it more difficult to satisfy the first four conditions necessary for successful collusion. Reaching consensus is harder with more parties involved. Detection is more difficult since price cutting by one small firm may be very difficult for some or all of the other firms to discern. Punishment is less likely to be effective for two reasons. First, a defecting firm is harder to deter because it has more to gain from cheating: its market-share gain during the period of initial defection is likely to be greater, and its loss from punishment smaller, the more firms were sharing in the collusive profits. Second, punishment may be more difficult to coordinate because of the free-rider problem. Inclusion is also harder due to free-rider problems, as each

⁵⁰ McCutchen (1997) even suggests that the Sherman Act may *help* firms collude, stating (p. 348): “Government policies that are designed to stop price fixing may benefit firms by making it worthwhile for them to meet to set up collusive agreements, while making it costly enough for them to avoid undesirable future negotiations.” However, this view is not supported by the *Sugar Institute* case reported in Genesove and Mullin (2001). In any event, evidence from the past decade shows clearly that large fines and treble damage awards in price fixing cases can and do impose substantial penalties on firms engaged in price-fixing, not the weaker sanctions necessary for McCutchen’s logic.

individual firm may believe that the others will coordinate, whether or not it participates.⁵¹ For precisely these reasons, one of the concerns underlying merger enforcement policy is that mergers between rivals that increase concentration can raise the likelihood that the remaining firms will coordinate after the merger.

Our simple supergame model with homogeneous goods and repeated price competition illustrates how symmetric collusive outcomes are more difficult to sustain when the number of suppliers is larger. We showed above that the monopoly price could be supported in a perfect equilibrium if and only if $\delta > \delta^* = 1 - 1/N$. The larger the number of firms, the larger is δ^* , meaning that the firms must be more patient to sustain the collusive outcome.

Asymmetries in market shares tend to make it more difficult to sustain collusion. For illustrative purposes, suppose that firm i has a market share of s_i .⁵² The condition for this firm to cooperate rather than defect is $s_i\pi(P)/(1-\delta) > \pi(P)$, which can be written as $s_i > 1-\delta$. This condition will be most difficult to meet for the firm with the smallest market share, since this firm has the greatest temptation to gain share before the other firms can respond and also the least profits to lose from punishment that renders all firms' profits equal to zero.⁵³ Defining s_{\min} as the market share of the smallest firm, we get $\delta^* = 1 - s_{\min} > 1 - 1/N$, so the firms must be more patient to sustain the collusive outcome.⁵⁴ The smallest firm plays the role of the maverick, that is, the firm most prone to defection from the collusive outcome.

While instructive, this simple model is unable to capture the other factors noted above, which tend to be even more important in practice. Of particular note is the temptation of one relatively small firm to decline to participate in the collusive arrangement or secretly to cut prices to serve, say 4% rather than 2% of the market. As long as price cuts by a small firm are less likely to be accurately observed or inferred by the other firms than are price cuts by larger firms, the presence of small firms that are capable of expanding is especially disruptive to effective collusion.

⁵¹ This point assumes that due to capacity limits, rising marginal cost, or other factors that some degree of collusion is possible even if some firms do not participate.

⁵² The analysis here is incomplete because it does not explain the underlying sources of the differences in market shares. We address that below when we consider cost asymmetries among the firms.

⁵³ The analysis assumes that the smallest firm is nevertheless able to capture the market; if, however, capacities are proportional to existing shares, the conclusion may not follow.

⁵⁴ Note that, when market shares are equal, $s_{\min} = 1/N$.

b) Cost Asymmetries

The calculation just given with s_{\min} is incomplete because it does not explain *why* the firms have different market shares. A common explanation is cost asymmetries, and these also make it more difficult for firms to sustain collusive outcomes.⁵⁵

Reaching consensus is clearly more difficult with cost asymmetries since there is less likely to be a focal point for pricing and since the firms may well disagree about the price they would like to see prevail. Furthermore, if collusion is to maximize potential industry profits, production efficiency requires that the low-cost firms be allocated a greater share of sales, but this may require contentious negotiations and/or side payments, thereby limiting somewhat the potential gains from collusion. These problems are exacerbated if cost information is private, since each firm may have an incentive to represent to the others that its costs are low in order to receive a higher allocation of output or other more favorable treatment.⁵⁶

The conventional wisdom states that enforcement of the collusive outcome is also more difficult with cost asymmetries. For example, Ivaldi et al. (2003) argue that cost asymmetries hinder collusion, stating (p. 36): “even if firms agree on a given collusive price, low-cost firms will again be more difficult to discipline, both because they might gain more from undercutting their rivals and because they have less to fear from a possible retaliation from high-cost firms.” Ivaldi et al. show in a simple duopoly example that a higher δ^* applies to the lower-cost firm than to the higher-cost firm, if the firms divide the market equally, because the lower-cost firm earns positive profits in the punishment phase. Assigning a larger share of the market to the lower-cost firm is one way to overcome this obstacle and restore the collusive outcome. However, allocating a lower share to the higher-cost firm necessarily makes it more attractive for that firm to deviate from collusion. Colluding firms thus face a tradeoff: lower-cost firms can be assigned larger market shares, which reduces their incentive to defect, but doing this increases the incentive of the higher-cost firms to defect.

These issues are explored in greater detail in Vasconcelos (2005), who studies repeated quantity competition among firms with heterogeneous quadratic cost functions, where firms differ in their ownership of an underlying asset that lowers the cost function.⁵⁷ He shows that, in the optimal collusive equilibrium, output is shifted away from the less efficient firms and towards the more efficient firms. In this equilibrium, the less efficient, smaller firms have the greatest incentive to depart from the collusive outcome, while the more efficient, larger firms have the greatest incentive to depart from the punishments specified by the equilibrium strategies. His results are relevant for the analysis of horizontal mergers since he shows how a merger affects the scope for collusion by changing not only the number of firms but also the distribution of holdings of the underlying asset and thus the distribution of costs among the firms.

⁵⁵ See Mason, Phillips, and Nowell (1992) for experimental results showing that cooperation is more likely in a duopoly if the firms have symmetric costs.

⁵⁶ Athey and Bagwell (2001, 2004) study repeated oligopoly with private cost information.

⁵⁷ See also Rothschild (1999).

Firms also may differ in their cost of capital and hence in the discount rate they use to compare current and future profits. A firm that is under financial pressure, for example, may have a high discount rate (low discount factor) and be especially tempted to defect. Harrington (1989) studied collusion among firms with different discount factors, showing again how market shares must be allocated to support a collusive outcome. As another example, a firm may believe it “deserves” a greater market share than it has historically enjoyed, perhaps because it believes it would greatly increase its market share under competitive conditions. Such maverick firms may be especially disruptive to collusive pricing.

c) Buyer Concentration and Auction Markets

Collusion is generally thought to be more difficult to sustain in markets where the buying side is highly concentrated. Apart from the fact that larger buyers may have a more credible threat to vertically integrate upstream than smaller buyers, buyers who purchase a large share of the output of the colluding firms can act strategically and internalize many of the benefits of disrupting collusion. For example, Snyder (1996) shows how a large buyer can strategically accumulate a backlog of unfilled orders to create a bulge in demand that can undermine or destabilize collusion. More generally, a large buyer can strategically create variations in demand over time. For example, by curtailing purchases in one period, the buyer may lead some or all of the suppliers to suspect that others have cheated on the pricing agreement.

Additional strategies are available for a buyer who is setting up the rules by which the suppliers will bid for business. Klemperer (2002) reports enormous variations in the prices received in auctions of third-generation mobile telephone licenses across different European countries, arguing that some auction designs facilitated collusion and thus led to far lower prices being paid for these licenses than was paid for other, comparable licenses. Marshall and Meurer (2004) discuss some of the unique issues that arise when considering collusion in a bidding context, including a discussion of spectrum and timber auctions, arguing that collusion is much more difficult in sealed-bid, first-price auctions than in oral ascending-bid auctions, a point proven more formally in Robinson (1985).

d) Collective Market Power Including Entry Barriers

If the firms have little collective market power, so they collectively face rather elastic demand for their products, their incentive to collude is correspondingly low. Collective market power may be small because the colluding firms are just a subset of the incumbent suppliers, because of low barriers to entry into the sale of the products the firms offer, or because the products they sell face competition from close substitutes sold by other firms. The smaller is the collective market power of the firms that are allegedly colluding, the closer will be the firms’ price to the competitive price, and the smaller the damages imposed on consumers by effective collusion.

e) Multi-Market Contact

Multi-market contact refers to situations where firms interact in more than one market at the same time. Much of the literature suggests that multi-market interaction tends to make it easier for the firms to sustain collusion. The standard reference here is Bernheim and Whinston (1990). Bernheim and Whinston first prove an irrelevance result: when identical firms with constant

marginal cost meet in identical markets, multi-market contact does not aid in sustaining collusion. While defection in one market can be punished in other markets, a firm can simply defect in all markets simultaneously. However, Bernheim and Whinston go on to show how multi-market contact can sustain collusion in many other settings. For example, multi-market contact can mute market level asymmetries, for example, if each firm has a major competitive advantage in one market (which could include one geographic area of a single product market). Suppose, for example, that Firm A is the leader in Market A and Firm B is the leader in Market B, but both firms compete in both markets. Firm A will be especially tempted to defect in Market B, where Firm A has a smaller share, but may be deterred if Firm B would respond in Market A. Mutual forbearance may well result. Furthermore, multi-market contact increases the frequency of interaction, permitting one firm to discipline another more rapidly than would otherwise be possible.

There is some evidence to support the proposition that multi-market contact makes it easier for firms to sustain collusive outcomes. In the airline industry, Evans and Kessides (1994) find that fares are higher on routes for which the carriers interact on multiple routes. In the mobile telephone industry, Parker and Röller (1997) find higher prices in markets where carriers have multi-market contact. Cramton and Schwartz (2000) look at signaling to support collusion in FCC spectrum auctions, where multiple auctions for licenses were conducted simultaneously.

The value of multi-market contact in sustaining collusive outcomes is less clear, however, once one accounts for the noisiness of the signals that the firms receive regarding possible defections by others. Green and Porter (1984) show that, with noisy signals, limited punishments are optimal; in their model, the firms revert to cooperation after a limited period of time. With multi-market contact, spreading punishment across markets may simply not be desirable, just as engaging in a longer price war, while feasible, may not be optimal in the Green and Porter model. After all, in models where punishments actually occur in equilibrium, stronger punishments are costly. This important idea is absent from the many supergame models in which punishment never takes place in equilibrium. Thus, in more realistic models in which defections and/or punishments actually occur, multi-market contact may have no effect on the ability of the firms to collude. This view is supported by the *Sugar Institute* case described by Genesove and Mullin (2001); the Sugar Institute was very careful to calibrate punishments to the violation and certainly did not employ the maximum possible punishment. Had they done so, the cartel would have collapsed early on. In fact, the Sugar Institute steered away from multi-market linkages, carefully limiting punishment to the same geographic region where the violation occurred.

5. Product Differentiation

The traditional view in antitrust circles has been that collusion is easier to sustain among firms selling homogeneous products rather than highly differentiated products. Reaching consensus should be easier when agreement only requires that one price, not many, be established. However, there is no compelling theoretical reason to believe that detection and punishment are more difficult if the products are more differentiated. On one hand, with highly differentiated products, a single firm that cuts its price is likely to gain relatively few sales, since many customers will still prefer the other brands. Therefore, defecting is less attractive. On the other hand, punishments are weaker, since price cuts by the other firms have a smaller effect on the

profits of the defecting firm if products are highly differentiated. Ross (1992) presents two models of oligopolistic supergames with differentiated products that capture these ambiguities.

Another reason that collusion is more difficult to maintain when products are differentiated is that dimensions of competition other than price and cost cutting can take center stage. Collusion along such dimensions as product design and marketing can be very difficult to establish and sustain. Even if an initial agreement is reached, the most tempting way to defect from a collusive agreement may be to improve one's product or to expand one's marketing budget rather than to cut one's price. The firms may find it very difficult to restraint competition on marketing because of the difficulty in drawing the line between permissible and impermissible marketing activities. Likewise, collusion on product design may be hard to sustain due to the difficulties of defining what types of product improvements are permissible and the fact that product improvements include an element of commitment, tempting firms to preempt their rivals to capture more market share on a sustained basis.

6. Capacity Constraints, Excess Capacity, and Investment in Capacity

In many industries, certainly including traditional manufacturing, capacity constraints are an important aspect of the competitive environment. In fact, capacity investment decisions can be the most important dimension along which competition occurs in the long run. We now address capacity decisions and their interaction with pricing and output decisions. We begin with a short-run analysis, which takes capacities as fixed, and then go on to the long-run analysis, which includes capacity investment decisions.

a) Collusion on Prices with Capacity Constraints

We have already observed that collusion on prices is easier to sustain if any single firm could only gain limited sales by cutting its price. As we noted, one reason a firm may not be able to increase its sales much by cutting price is that the firm may face capacity constraints.⁵⁸ Therefore, it would appear that collusion on prices is very easy to sustain if the firms all have little excess capacity.

This argument, however, is seriously incomplete. Most fundamentally, if all firms are producing at capacity, it is hard to say that they are effectively colluding on prices. Full-fledged price competition could not cause prices to be lower than the level at which demand and supply would be equated given full capacity utilization. Effective collusion on prices must, therefore, go hand-in-hand with some degree of output restriction, that is, excess capacity. With this clarification, one can ask how the presence of capacity constraints affects the analysis already provided in which such constraints were absent. Put differently, does the presence of excess capacity make it easier or more difficult to sustain collusion?

Following the literature, we frame this discussion in terms of firms that can produce at constant marginal cost up to some well-defined capacity level and not beyond that point (in the short run). More generally, one could study models in which each firm has a smoothly increasing marginal

⁵⁸ A defecting firm might be able to relax this constraint by building inventories in anticipation of cutting its price.

cost curve. The resulting analysis would be considerably more complex but lead to similar tradeoffs and conclusions.

The effects of symmetric capacity constraints on collusion are theoretically ambiguous. The greater is the excess capacity at each firm, the more each firm can gain by defecting. However, by the same token, greater excess capacity means that the other firms can expand output more to punish the defecting firm. In a price setting supergame with capacity constraints, Brock and Scheinkman (1985) show that collusion is more difficult to sustain in the presence of capacity constraints than in their complete absence, but the relationship between δ^* and the per-firm excess capacity at the monopoly price is not monotonic. Lambson (1987) generalizes these results to optimal cartel punishment strategies. Abreu (1986) obtains similar results for repeated quantity-setting games with capacity constraints.

Notwithstanding these theoretical ambiguities, in practice symmetric capacity constraints may well facilitate collusion, at least in comparison with a situation in which all firms can produce at constant marginal cost. After all, a far greater percentage expansion of output is likely to be needed for a lone defecting firm fully to benefit from price cutting than is needed for all of the firms to meet the expanded demand at the lower competitive price (especially if N is not very small). Plus, expansion for the latter purpose can take place over time.

Asymmetries *per se* in capacity constraints are likely to hinder collusion. More precisely, for a given level of total capacity, collusion is more difficult if capacity is distributed unevenly across the firms. If one firm has greater excess capacity, that firm has a greater incentive than others to cut its price, and its rivals have less of an ability to discipline that firm.⁵⁹

b) Capacity Investment Decisions

In the longer run, the firms can adjust their capacities. One can think of these capacity choices much like quantity choices and thus interpret the results from quantity-setting supergames as applying to capacity choices over time. This approach is most reasonable if capacities are relatively short-lived, so there is little commitment value associated with capacity, or if the market is growing, so that firms are routinely adding capacity. With this interpretation, the discount factor reflects the time over which one firm can observe other firms' capacity choices and respond with its own. Since it takes longer to change capacity than to change price, the discount factor relevant for capacity decisions is lower than that for pricing decisions, making collusion on capacities more difficult to sustain, *ceteris paribus*. On the other hand, the initial capacity expansion by the defecting firm can itself take time and may be difficult to hide, making it hard for one firm to gain much of an edge on its rivals before they are able to respond.

However, treating capacity decisions just like output decisions may fail to reflect accurately some of their distinctive aspects: capacity investments tend to be lumpy and involve significant sunk costs. The irreversible nature of capacity choices is emphasized by the literature on preemptive capacity investment, which predicts outcomes that are *more* competitive than the

⁵⁹ See Compte, Jenny, and Rey (2002), Davidson and Deneckere (1984, 1990), and Lambson (1994, 1995).

static Cournot equilibrium.⁶⁰ The commitment aspect of building capacity tends to make collusion on capacity more difficult: after one firm adds capacity, it may not be credible for the other firms to add capacity as well, or not as much as would be needed to deter the initial expansion.

Capacity choices can interact with pricing choices over time in complex ways. Benoit and Krishna (1987) show that firms will choose to build and maintain excess capacity to support a collusive pricing outcome. Davidson and Deneckere (1990) study a “semi-collusive” equilibrium in which the firms first pick capacities and then play a repeated pricing game, setting prices at the highest sustainable level.

7. Market Dynamics

a) Demand Growth, Demand Shocks, and Business Cycles

In the simple models of repeated pricing competition, demand growth makes collusion easier because a defecting firm sacrifices more in future profits in exchange for a short-term increase in its market share. Likewise, if demand is declining, defection is more tempting. One can easily incorporate these ideas into the simple model presented above by adding a market growth factor.

However, these results rely on several assumptions that may not be justified in the presence of market growth or decline. First, they assume that defection today will forever disrupt collusion and lead to highly competitive outcomes in perpetuity. We already observed that the firms will be tempted to renegotiate to avoid this unpleasant outcome. The incentive to renegotiate is greater if the market is growing. We have also emphasized that applying very strong punishments is not optimal in the presence of imperfect detection and that punishments proportional to the deviation are attractive. The lack of proportionality between today’s defection and perpetual punishment is even greater in the presence of growing demand. Second, the results ignore the possibility that growing demand will induce new firms to enter the market. The prospect of future entry makes it more tempting to defect in the present and less valuable to maintain cooperation among the current incumbents. Third, the simple model of repeated price-setting does not account for the fact that growing demand may tempt the firms to engage in preemptive capacity additions.

The logic of collusion implies that the temptation to defect depends upon the relative size of current versus expected future demand. This has implications for short-term demand shocks, which are distinct from secular growth or decline in demand. Rotemberg and Saloner (1986) study a model in which demand shocks are independently and identically distributed, so demand today conveys no information about demand in the future. A positive demand shock thus makes defection relatively more attractive. This same logic can be applied to collusion over the business cycle. Haltiwanger and Harrington (1991) show that collusion is more likely to break down during the portion of the business cycle during which demand is declining. Bagwell and Staiger (1997) generalize these results to a model in which demand alternatives stochastically between boom and recession phases. Porter (1983) and Ellison (1994) apply some of these ideas

⁶⁰ See the models of two-stage competition in Shapiro (1989) and the citations therein.

to the Joint Economic Committee, a railroad cartel from the 1880s. In a nice empirical application, Borenstein and Shepard (1996) find that retail gasoline margins are higher when future demand is expected to be higher or future costs are expected to be lower. Lastly, we stress that when demand is unpredictable, as in Green and Porter (1984), collusion is more difficult to sustain because the firms have greater difficulty distinguishing demand fluctuations from cheating on the collusive agreement.

b) Disruptive Innovation

The more likely it is that the market will experience a disruptive innovation, the harder it is to sustain collusion. To see this, suppose that each period there is some probability, ψ , that a major new technological innovation will be introduced into the market, disrupting the collusive agreement. (A similar analysis applies to other factors that might disrupt the agreement.) For example, a major innovation may disrupt the collusive agreement because it is introduced by a new entrant or because it introduces such a sharp asymmetry among the existing firms that cooperation is no longer sustainable. Suppose that the innovation ends the profit flows for the incumbent suppliers. Under these conditions, the payoff from defecting remains at $\pi(P)$ but the payoff from cooperating is reduced because future profits must also be discounted by the probability that disruption occurs. Formally, this is equivalent to changing the discount factor from δ to $\delta(1-\psi)$, making collusion more difficult to sustain.

c) Switching Costs, Network Effects, and Learning by Doing

Defection is more tempting if the defecting firm can gain a lasting advantage over its rivals, either in terms of market share or cost. With consumer switching costs, at least some of the customers gained today from a price cut will remain in the future even if prices fall once the defection has been observed. Capturing customers today has lasting value for the defecting firm: in models of competition with switching costs, a firm's installed base is a valuable asset, even if the firms compete vigorously to gain new customers. The logic here is similar to that of cutting price when customers' demand is high: the defecting firm captures more sales by cutting its price today. On the other hand, in the presence of customer switching costs it can be more difficult to attract customers in the first place.

Collusion also can be difficult to sustain in the presence of strong network effects, at least if the firms sell incompatible products. In the clearest case, where the market is bound to tip towards one product standard or another, collusion between incompatible products is difficult to maintain since the firm that is losing the standards battle may be very tempted to engage in price-cutting, or some other tactic, to avoid entering a downward spiral.

A similar dynamic arises in the presence of learning by doing. If learning is based on cumulative output, a firm that expands its production today will experience lower costs tomorrow, thereby gaining a lasting advantage. Due to the commitment and preemption aspects of higher current production, a firm that is more aggressive today captures more profits in the future, making collusion more difficult to sustain in the presence of strong learning-by-doing effects.

D. Agreements under Antitrust Law

1. On the Meaning of Agreement

As described briefly in Section A, there seems to be a contrast between the economic and legal approaches to the regulation of collusive behavior. Under the economic approach, one first attempts to determine the existence of collusion and the magnitude of its effects and then considers which if any remedies is appropriate. Under the legal approach taken by antitrust, the first step is the determination of whether there exists an agreement, and, if there is, certain legal sanctions apply: in the United States, these are treble damages to injured customers, criminal penalties on perpetrators including fines and imprisonment, and possibly injunctions against particular practices.

The extent to which these approaches diverge depends importantly on the legal concept of agreement. One standard definition—found in dictionaries and common usage in many contexts—is that an agreement signifies harmony of opinion or action.⁶¹ Under that straightforward notion, collusion seems nearly synonymous with agreement. Indeed, a typical dictionary definition of collusion is a secret agreement or cooperation, suggesting further that the terms have the same meaning.⁶²

It seems, however, from legal materials—court opinions, agency pronouncements, and commentary—that the law’s notion of agreement is different, in particular, narrower. Nevertheless, it has remained somewhat mysterious just what more is required. Return to the classic example of an undoubted agreement: the secret meeting in a smoke-filled room at which competing firms suggest prices to each other, settle on a particular price, and indicate their assent to adhere to that price. Suppose we remove the smoke from the room, and then the room itself—for example, the firms might use a conference call or e-mail (or, as in one antitrust case, entering fares and symbols on a common electronic airline reservation system). Now, let us dispense with the secrecy: perhaps the firms might speak to each other through sequential press conferences. At the conclusion of this sequence, we have a sort of behavior that is often observed and is generally considered to be legal (that is, not to constitute an agreement). But why? Which step has anything to do with whether or not the firms *agreed* to anything?⁶³ (As already mentioned, it is the economist’s term, collusion, not the legal term, agreement, that often denotes secrecy.)

⁶¹ The definitions throughout are taken from Merriam-Webster’s Collegiate Dictionary (10th ed. 1993), without quotation marks or ellipses. Sometimes other definitions are listed as well.

⁶² If the legal term “agreement” was interpreted to require secrecy, then the law would in essence offer a complete defense whenever price-fixers are willing to reveal their plans, which they would have every incentive to do if that insulated them from legal liability.

⁶³ It may matter for other reasons whether communications are public. For example, buyers may value having information sooner. (However, buyers do not value means of communication that make collusion against them possible, even if one consequence is that they learn of adjustments to collusive prices somewhat sooner.) In any event, it is not clear how this consideration bears on whether there exists an agreement.

As one reads legal statements on the subject, it appears that communication is central to the inquiry. Again resorting to common definitions, communication refers to a process by which information is exchanged between individuals through a common system of symbols, signs, or behavior. By that standard, press conferences surely involve communication. So does virtually any other means of effective collusion.

Consider another simple example. In a somewhat remote area, there are two retail gasoline stations located on opposite corners of an intersection. Each posts its price on huge signs readily visible from the road—and, of course, from the other station. The competitive price is \$2.00 and the monopoly price \$3.00. One can easily construct a sequence of interactions—wherein each station owner posts various prices, waits to see the other’s response, then adjusts his or her own price, and so forth. We would predict that, even if neither benefited from a formal course in game theory, they might readily settle on a price near \$3.00. The time during which a defector could reap profits without response might be a matter of minutes, not months. Hence, successful collusion seems quite likely.

The legal question is whether the two owners have “agreed” to price at \$3.00. Suppose, as suggested, that the legal system gives content to the term agreement by asking whether the parties communicated with each other. Well, they did not *speak* to each other; they may not even speak the same language. However, in the relevant sense, they did speak to each other in a common language, that of price. The absence of words may have slightly lengthened the time it took to settle (agree?) on the price of \$3.00. And, should one station cut its price to \$2.90 (in the absence of any change in market conditions, such as a drop in the price of fuel from refineries), the other station owner’s quick response, cutting its price, say, to \$2.80, will be pretty unambiguous; it will be understood as an invitation to raise prices, an invitation that would be accepted by posting a \$3.00 price.

Examples like these seem to suggest that there is little, if any, difference between the legal requirement of agreement and the economist’s notion of collusion. Yet it also seems that few think that this is actually the case. Surely, it is believed, the law requires more: more evidence of agreement, usually through more evidence of communications.⁶⁴ Yet, as should now be clear, it is hard to tell what more is being sought. It seems that some different sort of evidence is required, but evidence of what?

Some legal utterances distinguish between “express” agreements and “tacit” agreements. Tacit ordinarily means that the communication does not use words or speech (which, by contrast, is what is meant by express). By that definition, the press conferences, being conducted using words, would constitute express rather than tacit agreements, but the gasoline station owners, using signs, would not be express agreements—unless, of course, one pointed out that a sign

⁶⁴ The “statutory language [of Sherman Act Section 1] is broad enough . . . to encompass a purely tacit agreement to fix prices, that is, an agreement made without any actual communication among the parties to the agreement. . . . Nevertheless, it is generally believed . . . that an express, manifested agreement, and thus an agreement involving actual, verbalized communication, must be proved in order for a price-fixing conspiracy to be actionable under the Sherman Act.” *In re High Fructose Corn Syrup Antitrust Litigation*, 295 F.3d 651, 654 (7th Cir. 2002) (opinion by Judge Posner). See also the further discussion of this case in note 72 below.

showing “\$3.00” is functionally equivalent to a sign showing “three dollars,” the latter, containing words rather than numerals, constituting an express rather than tacit communication. Likewise, one could consider sign language, other hand signals, winks and nods, and so forth. Indeed, it is hard to believe that a sensible legal regime would make legality—and heavy consequences—turn on subtleties of modes of expression and taxonomic disputes over which constitute “expressions” or “communications.”⁶⁵

Official legal pronouncements, although sometimes seemingly clear, are not that helpful either. U.S. Supreme Court opinions include famous statements such as the following:⁶⁶ “[C]onscious parallelism’ has not yet read conspiracy out of the Sherman Act entirely.”⁶⁷ But this merely indicates that purely independent action—such as different gasoline stations raising their prices in parallel when the price of oil rises—does not constitute an agreement. Or consider: “The essential combination or conspiracy in violation of the Sherman Act may be found in a course of dealing or other circumstances as well as in an exchange of words. . . . [A conspiracy may be found where] the conspirators had a unity of purpose or a common design and understanding, or a meeting of the minds in an unlawful arrangement. . . .”⁶⁸ This (partially question-begging) expression aligns substantially with the idea that successful collusion is sufficient. More recently (and more commonly quoted in modern cases), the Supreme Court has stated that evidence must be presented “‘that tends to exclude the possibility’ that the alleged conspirators acted independently. . . [that is,] that the inference of conspiracy is reasonable in light of the competing inference[] of independent action. . . .”⁶⁹ Here, the interpretation depends on the meaning of “independent.” If taken to mean “without regard to others,” then collusive behavior is not independent action and thus is sufficient to trigger liability. Yet another pronouncement (in a more recent case, but not one directly addressed to the agreement question) is that “[t]acit collusion, sometimes called oligopolistic price coordination or conscious parallelism [is] not in itself unlawful.”⁷⁰ Tacit collusion, however, is undefined and is not generally understood to be the same as conscious parallelism. What all of these court decisions and most other statements

⁶⁵ Not only do legal authorities devote little attention to defining “agreement,” but when other terms like “express” are employed, these too are not elaborated. Furthermore, although one standard definition of express is to represent in words, other standard meanings include to make known (regardless of the mode), to reveal impulses artistically, and to represent by signs and symbols, which covers the full gamut, including presumably most meanings that many using the term “express” intend to exclude.

⁶⁶ As noted earlier, the European Union has a similar agreement requirement that likewise extends beyond formal contracts and is imprecise. In *Dyestuffs*, the European Court of Justice elaborated the concept of a concerted practice as “a form of co-ordination between undertakings which, without having reached the state where an agreement properly so called has been concluded, knowingly substitutes practical co-operation between them for the risk of competition.” *ICI Ltd. v. Commission*, Case 48/69 [1972] ECR 619, ¶64. Although some sort of contact between the parties seems to be required, the Commission seems inclined to find behavior illegal even when the contact is indirect. See Bellamy and Child (2001).

⁶⁷ *Theatre Enterprises v. Paramount Film Distributing Corp.*, 346 U.S. 537, 541 (1954).

⁶⁸ *American Tobacco Co. v. United States*, 328 U.S. 781, 809-10 (1946).

⁶⁹ *Matsushita Electric Industrial Co. v. Zenith Radio Corp.*, 475 U.S. 574, 588 (1986), quoting *Monsanto Co. v. Spray-Rite Service Corp.*, 465 U.S. 752, 764 (1984).

⁷⁰ *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.* 509 U.S. 209, 227 (1993).

have in common is that key terms are not defined, the subject is not directly discussed in any depth (that is, for more than a paragraph), and no rationale is offered for deeming one set of scenarios to be legal and another illegal.

A further important complication is that it is well accepted that, whatever is required to establish an agreement, it is allowable (and typical) for the demonstration to be indirect, through circumstantial evidence. So-called “smoking guns” are not required. For example, if the law demands proof of direct verbal communications on the specific price and pattern of punishment, it might be argued that near-simultaneous price increases, and then declines in response to defections, are evidence of such communications and hence sufficient to establish a violation. The implicit logic is, “How else could this behavior be explained?” This perplexing question and some of the earlier discussions on the possible meanings of agreement require further attention to the role of communications in the economic theory of collusive behavior.

2. Agreement, Economics of Collusion, and Communications

Suppose, as seems to be believed by most, that the legal requirement of an agreement is satisfied only by certain types of communication: perhaps verbal statements or close equivalents, sufficiently directed at competitors, that relate closely to pricing behavior, and that may be responded to reasonably promptly, precisely, and directly. What, then, is the relationship between these more explicit sorts of communication and the economic theory of collusion?

Reviewing the economic theory of collusion as summarized in Sections B and C, communication may be relevant at a number of points. First, the nature of communications may bear on the ease of reaching consensus. In the example with two gasoline stations, rather simple communications seem sufficient. But if there are more firms, greater heterogeneity (in costs, products, or other features), more uncertainty about buyer behavior, or other complicating factors, greater negotiation may be required, which in turn might be facilitated by more explicit (direct, head-to-head, simultaneous, prolonged) communications. This view is not entirely obvious, however, for if all parties knew that they were limited to a few rounds of simple price suggestions, after which they must have reached agreement, it is possible that agreements would be reached more quickly and with greater likelihood (although they perhaps would also be less durable, due to misunderstandings). While this discussion is largely outside simple models of repeated oligopoly, which typically ask whether a price P can be sustained, these questions are addressed in the literature on “cheap talk,” cited above.

Second, in the detection of cheaters, explicit, detailed communication might also be helpful. If firm A’s cheating is noticed by firm B, firm B could tell others. Compte (1998) and Kandori and Matsushima (1998) address this possibility in a model where firms observe and can communicate their private information about past play. Alternatively, if other firms suspect that firm A is cheating, discussions with firm A (perhaps supported by firm A presenting original invoices or other information) might help clear up the matter, avoiding price wars due to mistaken inferences. The Sugar Institute operated very much in this manner, as described by Genesove and Mullin (2001). Oligopoly theory is slowly moving more in the direction of modeling these types of issues, at least by exploring the role of communications about private cost information. See Athey and Bagwell (2001, 2004).

Third, punishment might better be coordinated with more explicit communication. Determining the magnitude of the price cut and its duration, perhaps focusing punishment when firms' product lines and regions of operations vary, and other aspects of strategy might be worked out better. As with reaching consensus, however, greater opportunity for detailed communication may be a double-edged sword. As noted, the opportunity for renegotiation can also undermine punishment. In any case, it is generally assumed in formal models that some particular punishment strategy has been chosen and will be pursued; the question explored is whether the strategy, if pursued, would deter cheating *ex ante*, or whether the strategy is credible, not what communications may be necessary to select or effectuate the strategy.

Fourth, inclusion might be enhanced through detailed negotiations. This consideration is based on similar reason to that of reaching consensus and is likewise outside standard formal analysis.

In all, there are many reasons to believe—and it generally is believed—that greater opportunity for freer, more detailed, explicit communication tends to facilitate collusion (although there are some countervailing factors). If this is indeed the case, it follows that it is more important to prohibit more explicit forms of communication. Still, the question remains, why not prohibit all communication? The answer must be that various forms of communication—such as making price information available to customers—serve other, legitimate purposes, and that less explicit communications—such as sharing aggregated and lagged sales information through a trade association—are more likely to promote socially valuable functions than to facilitate collusion. This statement, too, is not obvious, for many socially valuable functions, such as the setting of compatibility standards for emergent technologies or the sharing of information about industry conditions, require highly explicit communication. Furthermore, as the case of the two gasoline stations illustrates, in some instances facilitating collusion requires very little explicit communication.

Additionally, directing the legal inquiry at the nature of communications—which themselves often cannot be observed by the tribunal but must be inferred from circumstantial evidence—raises what might be called a paradox of proof. Suppose available evidence indicates that, in the situation under scrutiny, collusion is especially easy and the danger of supra-competitive pricing is accordingly very high. Moreover, evidence conclusively demonstrates that we have experienced a collusive outcome—at roughly the monopoly price—for years. Who wins? Arguably, the defendants. They could argue that, precisely because collusion is so easy, they were able to achieve monopolistic results—and, they gleefully concede, will be able to continue to do so for the foreseeable future—without any meetings in smoked-filled rooms, elaborate negotiations, and so forth. Just a few public pricing signals and they were off. Moreover, since all have taken courses in strategy at business school and all are advised by the leading consulting firms and their affiliated game theory experts, coordinating punishment with only minimal, indirect communications is a snap. Hence, the very strength of the evidence of the ease and success of collusion makes it implausible to infer that the defendant firms actually met and had long discussions about price-fixing.

Reflecting on this case and other possibilities, it would seem that the relationship between the ease of collusion and the likelihood that there were sufficiently explicit communications to trigger liability under the agreement requirement (whatever it turns out to be) is not monotonic. Put differently, we are asking just how should the factors listed above, which make it more or

less difficult to sustain collusive outcomes, be incorporated into a price-fixing case in which the existence of an agreement is proved through circumstantial evidence.

Beginning at one end of the spectrum, suppose that industry conditions are such that it is extremely difficult for the firms to sustain a collusive outcome, for example, because there are many firms, low entry barriers, price-cutting by one firm is very difficult for rivals to observe, and demand and cost are highly variable. Under these industry conditions, we would not expect the firms to have engaged in unobserved meetings that meet our explicit communication requirement simply because such meetings would likely be futile. Moreover, if it is nevertheless asserted that collusion occurred, we just will not believe that an effective price fixing agreement was reached. The evidence on pricing and cost could not have been certain, and any uncertainty is naturally resolved against an inference of collusion because it would be nearly impossible under the observed industry conditions.

Consider next an industry in which conditions are such that collusion is somewhat easier to sustain, perhaps because the industry is more concentrated, has moderate entry barriers, pricing is more transparent, and demand and cost are less volatile. These industry conditions make it more likely, but still far from inevitable, that a collusive outcome could arise. Under these circumstances, if collusion indeed seems to have occurred, its very difficulty (but not such high difficulty as to blend toward impossibility) suggests that explicit communications may well have been employed to carry it off.

Toward the opposite end of the spectrum, consider an industry in which the conditions are highly conducive to collusion: highly concentrated, no prospect of entry, transparent pricing, and stable demand and cost. Think of the two gas stations. We now have the case with which we began, presenting the paradox of proof: the very ease of collusion negates the inference that there must have been elaborate, explicit communications.

In sum, as industry conditions move from those that make collusion nearly impossible to those that make it incredibly easy, the inference that there must have been highly detailed communications first becomes stronger and then weaker. It is rather hard to say where on this continuum the maximum inference arises, or in what intermediate range some given proof standard is satisfied.⁷¹

How does this paradox of proof square with the law and what we observe in practice? U.S. courts typically insist upon the presentation of various so-called “plus factors.” Yet these factors are often little more than indicators that collusion rather than purely independent behavior is

⁷¹ The more one pushes the logic underlying the inference of agreement, the more complex it becomes. For example, in the region in which collusion is moderately difficult, a slight increase in the ease of collusion makes it more likely that collusion was attempted, which raises the likelihood of a given type of explicit communication, but, conditional on collusion having been attempted, reduces the likelihood that communication was more explicit because, by hypothesis, collusion is becoming easier. The depiction in the text, which assumes a single peak, may be overly simplistic. Moreover, one supposes that different industry conditions in different combinations that contribute to the ease or difficulty of collusion may have varying effects on the need for more explicit communication and the forms that it will take.

likely to have occurred.⁷² As just explained, such factors indeed favor the inference of an illegal agreement if but only if we are on the “difficult” side of the maximum, where additional evidence indicating the ease or benefits of collusion makes the likelihood of the requisite communications higher. On the other side of the maximum, they make collusion more likely but explicit communications less likely.

We are unaware of any cases (nor have we ever heard anyone suggest the existence of any cases) in what we are referring to as the paradox region, that is, past the peak, such that evidence that collusion is more feasible makes the inference of detailed communications less likely. How can one explain this one-sidedness? One possibility is that, even though the law has had this character for over half a century in the United States (and for shorter, although significant, periods in many other jurisdictions), no one has really understood the nature of the legal requirement.

Another possibility is that all cases are in fact at the difficult side of the maximum. That is, there are no industries where successful collusion is at all likely in the absence of highly explicit communications. Observe, however, that if this were true, the agreement requirement would be superfluous. That is, if there exists collusion, there must have occurred the requisite communications to trigger the agreement requirement. Were this always true, nothing would need to be proved beyond the mere existence of collusion. (This would suggest that Posner’s aforementioned prescription would be implied by existing law, and thus not constitute a significant departure from it.)

Yet another possibility is that there are cases past the maximum, in the paradox range, but defendants are reluctant to advance the argument that the proof against them implies the absence of any agreement and hence victory. The reason is that, in conceding that collusion is easy, likely, and probably in fact has occurred and will continue, they fear that they will hurt their case. Defendants may suffer in the determination of liability because, as a practical matter, a fact finder (whether a jury, judge, or expert tribunal) is more likely to condemn them if they in fact operate in a situation inherently conducive to collusive outcomes and are likely taking advantage of it. They win on the formal law but lose because they show themselves to be greedy and behaving in an antisocial manner. In that event, it may be that *de facto*, the greater the danger of

⁷² Some of the most common factors seem to go little beyond requiring interdependent rather than independent behavior. For example, prominent plus factors seem to include various sorts of evidence showing that the firms’ actions are “against self-interest” in the absence of collusion. For a survey and critical commentary, see Areeda and Hovenkamp (2003, vol. 6, 241-250). Courts also frequently rely on evidence that purports to directly indicate the existence of an agreement. For example, Judge Posner in *In re High Fructose Corn Syrup Antitrust Litigation*, 295 F.3d 651, 662 (7th Cir. 2002) offers, among others, the following quotations from the alleged conspirators as evidence of the existence of the requisite agreement: “We have an understanding within the industry not to undercut each other’s prices.” “[O]ur competitors are our friends. Our customers are the enemy.” A competitor’s president is called a “friendly competitor” and mention is made of an “understanding between the companies that . . . causes us to . . . make irrational decisions.” As the above discussion explains, however, there can exist such an “understanding” and firms can view competitors cooperatively as a result of education about collusion, good advice, common sense and experience, and open communications (such as the gas stations’ posting of prices), so it is difficult to discern in what sense more than the existence of consciously interdependent, collusive interaction is required.

collusion, the greater the likelihood of liability, without regard to any inference that does or does not follow about explicit communications and the satisfaction of the agreement requirement. Additionally, if there is a sufficient prospect that liability will be found, defendants may be worried about penalties. The more they argue that collusion is easy, the more plausible will be high estimates of overcharges (in amount and duration) and thus the greater will be fines and damage payments.

This discussion may raise more questions than it answers, but we believe that it is, ultimately, clarifying. The economic analysis of collusion, although quite complex, is at least fairly straightforward in stating the question it addresses and the motivation for the inquiry it undertakes. Upon examination, the same cannot be said about the law's requirement of an agreement and the role of industry conditions in inferring that such an agreement exists. We hope to have advanced understanding in two ways: by being more precise about what agreement might mean, and, for a given definition, by being more explicit about the relationship between the economics of collusion and whether such agreement requirement is satisfied.

Two additional observations about the interplay between the economics of collusion and antitrust law are in order. First, under antitrust law it is possible for there to be a violation even when it is clear that no successful collusion occurred. If competitors meet formally, enter into a written agreement, but ultimately fail miserably in executing it, most legal regimes would find a violation. If fines or damage awards were limited to a multiple of the overcharge, this finding would be moot. However, other sanctions may be employed; notably, those engaged in the attempt may be put in prison. It is sometimes efficient to punish unsuccessful attempts (especially when detection is difficult and limits on sanctions may make it impossible to punish violators sufficiently to achieve effective deterrence), and examining direct communication may help to identify unsuccessful attempts. Of course, evidence about pricing patterns of the sort that might be deciphered by economic experts may also aid in the task, especially if there were efforts to put the agreement into effect.

Second, separate from the agreement requirement, penalties may depend substantially on the extent and duration of overcharges. Undertaking these measurements requires expert economic analysis. The greatest difficulty, of course, is in determining what would have been the price but for the collusion. It is necessary both to specify conceptually the nature of the equilibrium that would otherwise have prevailed (perfect competition? monopolistic competition in price with differentiated products?) and to calculate just what price would have prevailed in that equilibrium.⁷³ This inquiry is very closely related (in some respects, identical) to that necessary to identify whether collusion existed in the first place.

E. Other Horizontal Arrangements

Our analysis has focused almost entirely on collusion that involves arrangements purely concerned with the fixing of prices. Simple price-fixing, in turn, is unambiguously—"per se"—

⁷³ If damages were based not on the overcharge times the quantity purchased, as is ordinarily the case, but instead or also on losses of consumer surplus regarding units not purchased, information on the entire relevant segment of the demand curve would be required.

illegal in the United States and subject to similar prohibitions elsewhere. There exists, however, a variety of horizontal entities—partnerships, trade associations, joint ventures, standard-setting bodies, to name a few—and such entities engage in myriad forms of conduct.

Certain horizontal arrangements can serve as substitutes for direct price-fixing. As noted, firms might agree to divide territories or customers, so as to eliminate competition. Although no particular price has been set, each firm is left to act as a monopolist with respect to its portion of the market, so the result is similar to that of a price-fixing cartel. The economic analysis is analogous: firms must be able to agree on the market allocation (instead of the price), cheating (selling to other firms' allotted customers) must be detectable and subject to effective punishment, firms with significant capacity need to be included in the agreement, and entry must be limited. Likewise, legal scrutiny tends to be similar: pure horizontal divisions of the market among competitors are also *per se* illegal in the United States.

Not all horizontal arrangements involve pure schemes to fix prices or divide the market (often called “naked restraints”). Nevertheless, many horizontal arrangements pose some risk to competition. Accordingly, antitrust laws need to draw distinctions. Under the law in the United States, this is done under the rubric of the “rule of reason”: reasonable schemes are permissible; unreasonable ones are prohibited. Obviously, this concept needs to be fleshed out and related to economic analysis, and this task will be our central focus in this Section. But first we will consider a particular class of horizontal arrangements that is closely related to our foregoing discussion of collusion and the legal prohibition on price-fixing.

1. Facilitating Practices

In our consideration in Section C of conditions bearing on the likelihood of successful collusion, we largely took such conditions to be exogenous. Some factors, however, are within the firms' control, individually or collectively. Antitrust scrutiny has focused primarily on the latter.⁷⁴ In this regard, two lines of attack must be distinguished. First, is horizontal agreement on some practice that facilitates collusion itself an illegal agreement in restraint of trade and thus an independent basis for liability? Most challenges, and our own discussion, emphasize this inquiry. Second, does the use of facilitating practices constitute evidence of the existence of an underlying agreement directly to fix prices? In some respects, the distinction may be immaterial, notably if both agreements on the facilitating practice itself and agreements on price-fixing are illegal and if the remedy is the same. (A remedial difference is that a facilitating practice might independently be enjoined.) Nevertheless, in the context of evaluating the evidence in a particular case, it clarifies thinking to keep this difference in mind.

⁷⁴ In *E.I. du Pont de Nemours & Co. v. Federal Trade Commission*, 729 F.2d 128 (2nd Cir. 1984), the FTC unsuccessfully challenged what it asserted to be facilitating practices that were unilaterally adopted (although employed by all four firms in the industry), claiming authority under Section 5 of the Federal Trade Commission Act, which does not require the existence of an agreement. To the extent that facilitating practices can only be challenged when their existence is attributable to an agreement, emphasis is placed on the issues considered in Section D, notably, under what circumstances an agreement can be inferred when multiple firms employ a facilitating practice.

An important facilitating practice that has long been the subject of antitrust regulation concerns information exchanges among competitors, sometimes in the context of trade association activity and other times conducted independently.⁷⁵ For example, in the *American Column & Lumber* case, a violation was found where firms exchanged information on prices in individual transactions and this information was subject to audit for accuracy.⁷⁶ Such information greatly eases the detection of cheaters, whereas such details do not have an obvious and substantial productive use. In *Container Corporation*, firms were deemed to have violated the antitrust laws when they called competitors to verify the accuracy of buyers' assertions of having been offered lower prices elsewhere.⁷⁷

Another class of interest involves firms' contracts with their customers. For example, a firm's use of a most-favored customer clause—under which it agrees to give all customers under contract the benefits of any price cut extended to a subsequent customer—may greatly reduce its incentive to defect from a collusive price since it must sacrifice profits on its existing customer base that was otherwise locked in for a period of time at a higher price. (This disincentive would be immaterial for an infinitesimal price cut, but if as previously discussed greater price cuts are necessary to attract substantial new business, the disincentive could be substantial.) Some firms employ price-matching (meeting competition) clauses, under which they promise to lower their price if the buyer can find a competitor that charges less. (Some clauses promise to equal the price, perhaps even retroactively, that is, on previous sales, and others promise to exceed the competitor's price reduction.) This arrangement deters other firms from lowering their prices. Moreover, it facilitates detection because buyers offered lower prices are more likely to reveal otherwise secret price cuts. Observe that under these arrangements buyers as a whole are disadvantaged—if effective, the market price is sustained at a higher level—but individual buyers are subject to the free-rider problem: each may well gain (if there is any chance that some seller will lower price), but its contribution to a higher market price will be negligible if it is a small purchaser.⁷⁸ These cross-currents are explored in the economic literature on most-favored customer clauses and meeting competition clauses.⁷⁹

Other types of practices are directed at coordination problems caused by product heterogeneity and competition along dimensions other than price. Quality or grading standards may promote uniformity—or at least reduce variety to manageable proportions—facilitating agreement on price. Agreements may limit credit (and other) terms, lest firms cheat on the price by offering favorable interest rates.

⁷⁵ As elsewhere, our discussion focuses on U.S. antitrust law. EU law also encompasses facilitating practices, including exchanges of detailed information among competitors in industries prone to collusion. See, for example, Bellamy and Child (2001, §4-042).

⁷⁶ *American Column & Lumber Co. v. United States*, 257 U.S. 377 (1921).

⁷⁷ *United States v. Container Corporation of America*, 393 U.S. 333 (1969).

⁷⁸ This type of free-riding problem also arises when a monopolist employs exclusive dealing provisions with its customers, as discussed in Section V below.

⁷⁹ See Cooper (1986), Edlin (1997), Edlin and Emch (1999), and Salop (1986). Borenstein (2004) applies this idea to price matching in the airline industry and the Airline Tariff Publishing case.

More broadly, any factor that may inhibit collusive pricing is potentially subject to firms' creativity in devising means of avoiding its detrimental effect. There are, of course, limits on what is feasible. Furthermore, to the extent that the use of facilitating practices itself requires collusion, firms must overcome any difficulties of coordination, detection, and enforcement with regard to the facilitating practices themselves, as illustrated by the *Sugar Institute* case. Some facilitating practices may be readily formulated and observed; others may be complex or hidden. Accordingly, the successful use of facilitating practices will vary greatly.

There remains another important consideration with many facilitating practices: they may have redeeming virtues. Some information exchange enhances planning. Forcing trading into formal markets (which was permitted in *Chicago Board of Trade*⁸⁰) produces benefits that flow from public prices. Contractual arrangements with buyers regarding the sellers' and competitors' prices reduce search costs. Exchange of cost information may enhance productive efficiency by shifting output to more efficient firms; see Shapiro (1986). Even what may seem literally to be price-fixing will often be efficient, such as when productive partnerships or joint ventures are formed and the resultant entity fixes a single price for its common product. Likewise, many other arrangements that may seem beneficial may also have effects on the feasibility of collusion. Accordingly, it is necessary to formulate a means of balancing the costs and benefits, which is the subject of the next subsection.

2. Rule of Reason

In the United States, the "rule of reason" was formally announced nearly a century ago, in the monopolization case of *Standard Oil*.⁸¹ Shortly thereafter, it was given more content in *Chicago Board of Trade* in language that is routinely quoted (or paraphrased) to this day: "The true test of legality is whether the restraint imposed is such as merely regulates and perhaps thereby promotes competition or whether it is such as may suppress or even destroy competition."⁸² Although more specific than the almost-completely question-begging inquiry into "reasonableness," the meaning of this test is hardly self-evident. Just what is meant by "competition"? Is it valued purely as a means or as an end unto itself? It is useful to begin with a few modern invocations of the rule of reason's promoting competition test, followed by some reflection on the broader question of interpretation and its relationship to economic analysis.⁸³

⁸⁰ Board of Trade of City of Chicago v. United States, 246 U.S. 231 (1918).

⁸¹ Standard Oil Co. of New Jersey v. United States, 221 U.S. 1 (1911).

⁸² Board of Trade of City of Chicago v. United States, 246 U.S. 231, 238 (1918). The Court continued: "To determine that question the court must ordinarily consider the facts peculiar to the business to which the restraint is applied; its condition before and after the restraint was imposed; the nature of the restraint and its effect, actual or probable. The history of the restraint, the evil believed to exist, the reason for adopting the particular remedy, the purpose or end sought to be attained, are all relevant facts. This is not because a good intention will save an otherwise objectionable regulation or the reverse; but because knowledge of intent may help the court to interpret facts and to predict consequences."

⁸³ In the European Union, conduct may be deemed exempt from the prohibition in Article 81(1) on anticompetitive agreements if it meets certain criteria in Article 81(3) that bear resemblance to the rule of reason in the United States. There are both block (general) exemptions and those granted individually. To enhance clarity, the

In *National Society of Professional Engineers*, the society had an ethics rule prohibiting engineers from bargaining about price until after they were selected for a project.⁸⁴ The proffered justification was that otherwise customers might be induced to focus excessively on the price of professional services at the expense of concerns about quality and safety. The Supreme Court found a violation. Safety was not deemed unimportant, but rather something that ultimately was for customers to decide. They could employ the society's approach if they wished, but the society could not impose this choice on all customers. Competition meant free and open choice, not one side of the market dictating terms to the other.

In *Indiana Federation of Dentists*, cost-conscious insurance companies employed an internal procedure for reviewing submissions for reimbursements.⁸⁵ The dentists objected (to non-dentists passing professional judgment on their work, so they claimed) and agreed as a group not to supply the necessary documentation. They too lost. Once again, it was for the customer—or, in essence, the customer's agent, the insurance company—to make whatever judgments it wished. Any individual dentist was free not to deal with any insurer if the dentist thought the insurer's practices inappropriate (or for any other reason or for no particular reason), but dentists could not agree, as a group, to impose the judgment.

Consider also *National Collegiate Athletic Association*, involving agreements among universities regarding college football.⁸⁶ The Supreme Court was not bothered by their agreements on rules of the game (size of playing field, scoring, and so forth)—rules that were not challenged—but did find their restrictions on schools selling television rights independently of the Association's scheme to constitute a violation.⁸⁷

For the most part, cases such as these seem to view competition as a process. The view seems to be that competition consists of buyers and sellers each deciding for themselves—or, more precisely, in individual buyer-seller pairs—with whom they will deal and on what terms. Independent decisions are a central feature of competition, whereas groups (typically of sellers) who attempt to impose some regime regarding the proper terms of dealing are subverting the process. They may or may not be right, but that is not the question. Put another way, what is right is essentially taken to be whatever is the outcome of the competitive process, much like how one accepts the equilibrium price in a competitive market as “reasonable.”

Perhaps competition is viewed as good in itself. Or instead the view may be that competition is valued for its results, whether those understood by economists, in terms of allocative efficiency,

Commission has issued Guidelines on the Applicability of Article 81 of the EC Treaty to Horizontal Cooperative Agreements (European Union 2001).

⁸⁴ *National Society of Professional Engineers v. United States*, 435 U.S. 679 (1978).

⁸⁵ *Federal Trade Commission v. Indiana Federation of Dentists*, 476 U.S. 447 (1986).

⁸⁶ *National Collegiate Athletic Association v. Board of Regents of the University of Oklahoma*, 468 U.S. 85 (1984).

⁸⁷ This case also nicely illustrates what is sometimes referred to as the “ancillary restraints” doctrine. Namely, an anticompetitive restraint is not deemed permissible merely because it is associated with an otherwise legitimate venture; however, the restraint may well be allowed if it is reasonably necessary to accomplish the legitimate objectives of the venture.

or other notions concerning freedom of choice. Under this second, instrumental view of competition, the antitrust laws are nevertheless interpreted to relate only to the process: perhaps the integrity of the competitive process is much easier to assess than the outcome of that process, and benefits are assumed to flow as long as competition is assured. In cases where it is alleged that the competitive process is not providing the expected benefits, courts in the United States repeatedly state that the appropriate remedy is to seek legislative or regulatory action. Even when there are market imperfections, it is plausible to distrust the collective schemes of self-interested market actors, schemes that they allege to be correctives in the public interest—Adam Smith’s warning being apropos.

This process view, however, is problematic. Although economists routinely use the term “competition,” it does not readily bear the weight it must in judging industry practices. Is the formation of a joint venture between two firms that might otherwise compete with each other, although less effectively or with a somewhat different product, an enhancement to or detraction from competition? How about a partnership or a horizontal merger? What of curing a market failure? Even if the result of coordinated action is unambiguously more efficient, is it more competitive? Does the competitive process include competition among institutional forms, including various forms of cooperation among groups of firms that operate in the same industry? More broadly, when the conditions for perfect, textbook competition fail (that is, pretty much always), is there an unambiguous way to describe one or another arrangement or outcome as more competitive?⁸⁸

Economists do not traditionally answer such questions. Instead, they undertake positive analysis of behavior and outcomes under various market arrangements. For normative purposes, the ordinary metric is welfare, or efficiency, or perhaps utility to each party or class of parties, not the degree of competition according to some competition index. Yet, if the rule of reason is legally defined in terms of competition itself—that which promotes competition is legal, that which suppresses competition is illegal, end of story—then economics cannot directly address the legal test.

As it turns out, no matter how often the promote-versus-suppress-competition test is invoked, it is not adhered to uniformly, and legal authorities seem to depart from it fairly readily in many of the cases in which its application seems problematic. As noted, in *National Collegiate Athletic Association*, the Supreme Court finds horizontal agreement on rules of the game to be unproblematic. There is a sense in which this flexibility may have benefited from a fortuitous play on words, in that such rules were seen as creating competition—sports competition, that is. But even the case that first announced the now-canonical language on competition, *Chicago*

⁸⁸ As with “agreement,” little aid comes from standard definitions. Competition is ordinarily taken to mean the act or process of competing, rivalry, or specifically the effort of parties to secure business of a third party. Under that rubric, even a simple partnership of two individuals who otherwise might produce (however inefficiently) on their own can readily be seen as “anticompetitive.” This definition is reasonably clear, but as will be discussed it is one that antitrust tribunals often disregard, and with good reason.

Board of Trade, was one that condoned restrictions on individual players' freedom of action to produce a greater good, trading in the public market.⁸⁹

Other modern cases reinforce a more complex interpretation. Notably, in *Broadcast Music*, there were two large entities (BMI and ASCAP) that, between them, licensed the rights to nearly all domestic copyrighted music to various users (for example, radio and television stations).⁹⁰ The entities set a single fee for block licenses, which was challenged, among other reasons, as constituting illegal price-fixing. And in fact, each of these two entities did set single prices for bundles of millions of musical compositions that otherwise might be priced independently. Yet, economies of scale in contracting and copyright enforcement (that is, monitoring the illegal use of the entities' portfolios of music by unlicensed parties) induced the Court to find no violation. The result in the market was nothing like atomistic competition under which individual composers paired voluntarily with individual buyers, an alternative the Court found to be cumbersome. Instead, huge collections of otherwise-competitors used a sales agent to dictate price and other terms of dealing. Supposing one accepts that the permitted arrangements were on-balance desirable on efficiency grounds, there remains the question of whether the arrangement involved more or less "competition."

One device employed in *Broadcast Music* and in some other cases is to treat the venture as a single entity: once viewed in this manner, there is no longer a horizontal agreement and thus no violation of Sherman Act Section 1. Looking ahead to Section IV, horizontal mergers are not themselves viewed as price-fixing cartels—even though the merged firms presumably fix a common price—but rather as single entities. In such cases, however, there remains the question whether the agreement *creating* what is subsequently viewed as a single entity constitutes a violation. *Carte blanche* would authorize formal cartels, say, incorporated as a single firm. Of course, jurisdictions do not freely permit formal cartel arrangements or horizontal mergers. Nor do they automatically approve even loose trade associations, if for example member firms engage in information exchanges of a sort that facilitate collusion and generate little offsetting benefit. That is, when trade association activity has been challenged successfully, no single-entity defense has been recognized. Thus, the single-entity characterization is more of a conclusion than a reason to decide one way or the other.

What, then, is the underlying meaning of the rule of reason? On one hand, antitrust law does not insist on pure atomistic competition, prohibiting all combinations from small partnerships to trade associations to joint ventures to mergers. On the other hand, horizontal arrangements are not freely permitted. Instead, they are subject to some sort of balancing test, whether under the rule of reason in the United States or under other rubrics elsewhere. When the arrangement looks like little more than a pure interference with ordinary competition, it is likely to be condemned with little further inquiry. In many cases in the United States, for example, market power need not be demonstrated, and adverse effects need not be proved (although for an award

⁸⁹ Our point is not to agree with the analysis in *Chicago Board of Trade*, which was problematic in a number of respects, but rather to indicate that the pure, atomistic, hands-off process view of competition was never the complete story.

⁹⁰ *Broadcast Music, Inc. v. Columbia Broadcasting System, Inc.*, 441 U.S. 1 (1979).

of damages these considerations will be important). Examples include *National Society of Professional Engineers*, *Indiana Federation of Dentists*, and the television marketing restrictions in *National Collegiate Athletic Association*—in addition to naked price-fixing and related practices. When, however, there appear to be benefits—from combining production, conducting research, setting standards, or otherwise—condemnation is not guaranteed, as demonstrated by *Broadcast Music*. And some horizontal arrangements, like partnerships and mergers that do not produce substantial market power, are routinely allowed.

The primary area of ambiguity concerns the many practices that fall in between the extremes. Economists can analyze the arrangements' effects and assess their efficiency. But how do such assessments relate to the legal test? Most modern antitrust rule-makers and adjudicators seem to pay substantial attention to economic considerations, at least in many settings. But under formulations like the rule of reason, the conception of reasonableness—whether or not concretized as a determination of promotion versus suppression of competition—is not well specified. We know from cases like *Broadcast Music* that pros and cons will sometimes be balanced, but what counts as a benefit or cost of an arrangement, what metric is employed for measurement and conversion to a common denominator (if this is done at all), and what is the ultimate decision rule remain somewhat a mystery.⁹¹ A purely economic criterion has not been explicitly embraced; nor has it been rejected.⁹²

F. Antitrust Enforcement

We close this section by commenting briefly on some of the law and economics issues that arise in antitrust enforcement.⁹³

1. Impact of Antitrust Enforcement on Oligopolistic Behavior

The workhorse model of oligopoly used to study collusion, namely the model of repeated price- or quantity-setting, does not explicitly include any antitrust enforcement. At first blush, this seems rather peculiar, at least from the perspective of law and economics. However, if this basic model captures conduct that is believed by the parties to be beyond the reach of the antitrust

⁹¹ “Courts sometimes describe their task under the rule of reason as one of ‘balancing’ potential harms against likely gains or defenses. But balancing implies that one places some measurably quantity of something on one side of the scale, a quantity of something else on the other side, and determines which side outweighs the other. The set of rough judgments we make in antitrust litigation does not even come close to this ‘balancing’ metaphor. Indeed, most courts do not define a unit of measurement in which the quantities to be balanced can be measured. Assuming the relevant unit is dollars, one would need to place at least a rough dollar estimate on the dangers to competition . . . and a similar estimate on likely cost savings, output increases, or other benefits. To the best of our knowledge, this has never been done in any antitrust case.” Hovenkamp (1998, vol. 11, 301-302). Hovenkamp, it should be noted, does not offer this depiction as a criticism. Instead, he sees such balancing as beyond the institutional competence of courts and believes that in practice they employ a structured sequence of (essentially dichotomous) inquiries that usually enables them to resolve cases one way or the other without ever having to balance costs and benefits.

⁹² One question of particular interest is, supposing that the criterion is economic, whether it involves efficiency as a whole or only consumers' surplus. Compare our discussion of this issue in the context of horizontal mergers, in subsection IV.D.3.

⁹³ For discussion of additional issues, see, for example, Posner (2001).

laws—repeated price-setting without any other communications—then the omission is justified. This is another reminder that economic theory may be most relevant in determining the existence of price-fixing when it helps us understand whether additional conduct, such as communications or facilitating practices, significantly increases the likelihood that a collusive outcome will occur.

In contrast, wherever antitrust law is applicable, it is important to consider the influence of expected sanctions on firms' behavior. Harrington (2004a, 2004b, 2005) introduces enforcement policy into oligopoly supergames. He posits that a newly formed cartel will be more likely to attract the attention of antitrust enforcers (perhaps based on complaints by customers) if it rapidly raises price from the competitive level to the cartel level. He shows how the price path adopted by the cartel and the steady-state cartel price are affected by antitrust enforcement. He also studies the relationship between damages rules in price-fixing cases and cartel pricing. In the process, he identifies some complex and even perverse effects of antitrust enforcement on cartel pricing.

2. Determinants of the Effectiveness of Antitrust Enforcement

A number of other aspects of antitrust enforcement have recently been illuminated by economic analysis. One increasingly active approach to enforcement is the government's attempt to strategically induce some colluding firms to turn on their peers. Enhanced lenience toward cooperating firms (as well as increased international cooperation) led to successful prosecutions in a series of major international price-fixing cases during the 1990s. Harrington (2006a) discusses the impact of corporate leniency programs on collusion. See also Motta and Polo (2003) and Motta (2004, p. 194) on the European Commission's newly adopted leniency policy, and Litan and Shapiro's (2002) discussion cartel enforcement during the 1990s.

Another important enforcement supplement that is particularly important in the United States involves private lawsuits for (treble) damages. When the Department of Justice brings a price-fixing case, there typically are immediate follow-on private actions brought by parties claiming to have been overcharged. Frequently, these cases are brought as class actions, and many have resulted in large payments. Although only direct purchasers can claim damages under U.S. federal antitrust laws, many states allow indirect purchasers to recover damages as well. In all of these settings, economists are relied upon to estimate damages for overcharges. As previously noted, the challenge they confront—determining what prices would have existed but for the illegal collusion—is closely related to the underlying analysis of collusive behavior.

There is also a growing economic literature on cartel detection—addressing what patterns of pricing, or bidding, are indicative of collusion—that is important for setting enforcement priorities, determining liability, and assessing damages. Porter (2005), Harrington (2007), and Whinston (2006, pp. 38-52) provide highly informative recent surveys on cartel detection. Bajari and Summers (2002) discuss the detection of collusion among bidders in an auction setting.

IV. Horizontal Mergers

The primary concern about horizontal mergers—that is, mergers between direct competitors—is that they may lead to anticompetitive price increases, either because the merged entity on its own

will find it profitable to raise prices from pre-merger levels (so-called unilateral effects) or because the increase in concentration enhances the prospects for successful collusion (coordinated effects).⁹⁴ Accordingly, we begin by offering an economic analysis of these possibilities, drawing on our analysis in Sections II and III. Next, we briefly review empirical evidence on the actual effects of horizontal mergers.⁹⁵

Antitrust enforcement plays an active role with regard to horizontal mergers because in the United States nontrivial mergers must be reviewed by one of the two federal authorities with overlapping jurisdiction, the Antitrust Division of the Department of Justice (DOJ) and the Federal Trade Commission (FTC). (Similar merger review takes place in other jurisdictions, such as the European Union.) These reviews are governed by the Horizontal Merger Guidelines, the current version of which was (mostly) promulgated in 1992 by the DOJ and FTC.⁹⁶ We describe the pertinent procedures, from initial filing to the agencies' analysis to court challenges and remedies. Although our focus is on the Merger Guidelines due to their current centrality, we also discuss the pertinent antitrust statutes and the evolution of horizontal merger doctrine in the courts. We pay particular attention to the role of prospective merger synergies, usually referred to in antitrust discussions as merger efficiencies. These benefits are important in determining the threshold of anticompetitive effects that must be present to challenge a merger—that is, the law implicitly presumes mergers to be advantageous to some degree—and also in offering a possible affirmative defense to a merger that otherwise would be prohibited. Furthermore, in assessing the role of efficiencies in justifying horizontal mergers, it is necessary to specify more precisely the goals of the antitrust laws, in particular, whether the objective is to maximize total economic welfare or instead just consumer surplus.

Finally, we consider in greater depth the economics underlying the analysis dictated by the Merger Guidelines, particularly with regard to market definition, relating the Guidelines approach to the economic analysis of market power presented in Section II. In this regard, we also discuss the growing body of empirical methods for predicting the effects of particular horizontal mergers. Part of the challenge is theoretical: given that there are a number of theories of oligopoly, with rather different predictions, which one should be used in a given merger? Presumably, the one that best fits the facts of that merger. But all of these theories are highly simplified in comparison with the inevitable complexity of real world competition, so picking the most suitable model of oligopoly is far from straightforward.

⁹⁴ A price increase often serves as a proxy for other possible anticompetitive effects, such as a reduction in product quality or service or a decrease in the pace of innovation.

⁹⁵ A subject related to horizontal mergers that we do not consider here is the tendency of partial cross-ownership to soften competition and thus increase price. See Bresnahan and Salop (1986), Reynolds and Snapp (1986), Farrell and Shapiro (1990a), O'Brien and Salop (2000), and Gilo, Moshe, and Spiegel (2006).

⁹⁶ U.S. Department of Justice and the Federal Trade Commission, Horizontal Merger Guidelines (April 2, 1992) (as revised April 8, 1997 with respect to Section 4, relating to Efficiencies). See also the enforcement agencies' detailed commentary on the guidelines. DOJ and FTC (2006).

A. Oligopoly Theory and Unilateral Competitive Effects

The basic idea underlying theories of unilateral effects is that the merged firm will have an incentive to raise its price(s), in comparison with the pre-merger price(s), because of the elimination of direct competition between the two firms that have merged. The examination of specific oligopoly models makes it possible to quantify the effects, which is important for merger enforcement. First, quantification can help to identify the mergers that are most likely to have significant price effects and thus cause significant harm to consumers. These are the mergers that presumably warrant further scrutiny, if not prohibition. Second, quantification allows us to estimate the merger efficiencies necessary to offset the loss of competition and thereby allow the merger to pass muster according to the consumer surplus or total welfare standard.

1. Cournot Model with Homogeneous Products

We begin by studying the effects of mergers in the Cournot oligopoly model described in the subsection II.C.1. The Cournot model seems like a good starting place since it generates a number of sensible predictions relating market structure to the equilibrium outcome. In

particular, we derived equation (3) $\frac{P - MC_i}{P} = \frac{S_i}{|\varepsilon_D|}$ that relates a firm's price-cost margin to its

market share and the market elasticity of demand. In the special case with constant and equal marginal costs, each firm has a market share of $1/N$, and the Cournot model predicts that the margin of each firm will be given by $(P - MC)/P = 1/N|\varepsilon_D|$. We also derived an expression for

the industry-wide average, output-weighted, price-cost margin, that is, $PCM \equiv \sum_{i=1}^N S_i \frac{P - MC_i}{P}$,

namely expression (4): $PCM = \frac{1}{|\varepsilon_D|} \sum_{i=1}^N S_i^2 = \frac{H}{|\varepsilon_D|}$, where, recall, $H \equiv \sum S_i^2$ is the Herfindahl-

Hirschman Index (HHI) of market concentration.

The idea that a firm with a large share will have more market power, and thus will charge a higher price (but still less than the monopoly price) has been very influential in horizontal merger enforcement. So has the idea that margins are higher in more concentrated industries. In fact, based partially on the expression for the PCM, the Merger Guidelines measure market concentration using the HHI. At the same time, it is recognized that the margins of all firms in a given market are lower if the elasticity of demand in that market as a whole is large, a subject to which we will return in discussing market definition in Section E below.

While all of these expressions accurately characterize the Cournot equilibrium, none of them actually tells us what happens to price, consumer surplus, profits, or total welfare as a result of a merger between two firms in a Cournot oligopoly. To answer those questions, which are central to the analysis of horizontal mergers, it is necessary to compare the Cournot equilibria before and after the merger and, in particular, to specify what is involved when two formerly independent firms become one.

Salant et al. (1983) address this question, emphasizing the peculiar result that mergers in a Cournot oligopoly can be unprofitable. The reduction in the number of firms raises price.

Initially, the merging firms reduce their output because they internalize more of the effect of their output on price than they did previously. In turn, non-merging firms raise output somewhat, leading the merging firms to cut output further. At the new equilibrium, price is higher. (Indeed, this is why price is ordinarily higher in Cournot equilibrium when there are fewer firms.) But the merged firm's combined share of total industry profits is lower; after all, other firms' quantities rise and the output of the merging firms falls. Salant et al. focus on the symmetric case with constant marginal costs; in this setting, a merger of any number of firms is equivalent to all but one of the merging firms shutting down. As a result, the cost due to the smaller profit share will exceed the benefit from a higher industry price unless, in their example, the merging firms constitute 80% or more of the industry! In this simple model, a merger does not lead to a "stronger" firm in any sense—as noted, it is as if the acquired firm simply exits. If this story depicts how mergers work, few mergers (short of mergers to monopoly) would be observed. Accordingly, a theory that plausibly explains mergers that actually occur requires that the merging firms own assets that can be usefully combined in some way.

Perry and Porter (1985) pursue this point using a model in which each firm owns a certain amount of capital. In their model, each firm's marginal cost increases linearly with that firm's output, and the slope of the marginal cost curve is lower, the larger is the firm's capital stock. Thus, firms that own more capital are larger in the resulting Cournot equilibrium. Perry and Porter assume that when two firms merge, the merged entity owns their combined capital stock and thus has a lower marginal cost curve than either of the constituent firms. In addition, since the marginal cost of each rival firm rises with its output, the ability of rival firms to expand in response to the merger is not as great as in the prior example in which marginal cost is constant. As a result, horizontal mergers are much more likely to be profitable in this model. Levin (1990) generalizes the Salant et al. model in a different direction by allowing the firms to differ in their (constant) marginal costs.⁹⁷ McAfee and Williams (1992) further explore models with quadratic cost functions where the marginal cost of a firm is proportional to the ratio of its output to its capital stock, showing how the magnitude of the price increase resulting from a merger depends on the capital stocks of the merging and non-merging firms.

Farrell and Shapiro (1990b) significantly generalize these results and provide an analysis of the price and welfare effects of horizontal mergers in Cournot oligopoly. They start with a Cournot equilibrium among N firms, where the cost function of firm i is given, as before, by $C_i(X_i)$. A merger in Cournot oligopoly can be modeled as the replacement of two existing firms with cost functions $C_1(X_1)$ and $C_2(X_2)$ by a single merged firm with its own, new cost function, $C_{12}(X_{12})$.

Farrell and Shapiro say that a merger generates no synergies if the merger simply allows the merging firms to rationalize output between their existing operations or facilities, that is, if $C_{12}(X_{12}) = \min_{X_1, X_2} [C_1(X_1) + C_2(X_2)]$ subject to $X_1 + X_2 = X_{12}$. Define the pre-merger outputs of

⁹⁷ Levin also allows the merged firm to behave other than as a Cournot oligopolist, for example, as a Stackelberg leader.

the two merging firms as \bar{X}_1 and \bar{X}_2 , and the pre-merger price as \bar{P} . Label the two merging firms so that firm 2's pre-merger output is at least as large as firm 1's pre-merger output, $\bar{X}_2 \geq \bar{X}_1$. Using the pre-merger Cournot equilibrium relationship (3), $\frac{\bar{P} - MC_i}{\bar{P}} = \frac{S_i}{|\varepsilon_D|}$, we know that larger firms have higher markups, so firm 1's marginal cost in the pre-merger equilibrium, $\overline{MC}_1 = MC_1(\bar{X}_1)$ is at least as large as firm 2's, $\overline{MC}_2 = MC_2(\bar{X}_2)$. Denote the merged firm's marginal cost at the combined output by $\overline{MC}_{12} = MC_{12}(\bar{X}_1 + \bar{X}_2)$.

Using this framework, Farrell and Shapiro prove generally the important result that mergers generating no synergies raise price. Without synergies, the merged firm's ability to rationalize production between its existing operations (by equating the marginal cost of production in the two operations) is not sufficient to offset the incentive to raise price that results from combining the ownership interests of the two operations.

Farrell and Shapiro also ask about the magnitude of synergies necessary for a horizontal merger to lead to a reduction rather than an increase in price. This is an important question in practice because, as discussed in subsection D.3, mergers tend to be judged based on their impact on consumers. Farrell and Shapiro provide a very general necessary and sufficient condition: a merger reduces price if and only if $\overline{MC}_2 - \overline{MC}_{12} > \bar{P} - \overline{MC}_1$. That is, the merger will reduce price if and only if the marginal cost of the merged firm (at the pre-merger combined output) is less than the marginal cost of the more efficient firm (at its own pre-merger output) by an amount that exceeds the difference between the price and the marginal cost of the smaller, less efficient firm prior to the merger. This inequality can be expressed in proportion to the pre-merger price as

$$\frac{\overline{MC}_2 - \overline{MC}_{12}}{\bar{P}} > \frac{\bar{P} - \overline{MC}_1}{\bar{P}} = \frac{S_1}{|\varepsilon_D|} \quad (6)$$

where we have added the pre-merger relationship between firm 1's margin and its share. This is a very demanding condition in an industry with moderate to large pre-merger margins. For example, consider a Cournot industry in which the market elasticity of demand at the pre-merger price is $\varepsilon_D = -1.0$, normalize the pre-merger price at $\bar{P} = 100$, and suppose that the pre-merger market shares of the two firms are 10% and 30%, so $S_1 = 0.1$ and $S_2 = 0.3$. Using the pre-merger Cournot equilibrium conditions, the pre-merger marginal costs of the two firms must be 90 and 70 respectively. The inequality above tells us that the merger will lower price if and only if the marginal cost of the merged firm, at the combined output of two merging firms, is less than 60.

Using these general results, Froeb and Werden (1998) provide calculations that relate the required magnitude of the synergies to the pre-merger shares of the merging firms. In the symmetric case, they show that the proportionate reduction in marginal cost necessary for price not to rise is equal to $S / (|\varepsilon_D| - S)$, where S is the pre-merger market share of each merging firm.

Analyzing the welfare impact of such mergers is more complex, in part because welfare effects depend heavily on the cost function of the merged entity in comparison with the cost functions of the two constituent firms, which captures any synergies resulting from the merger. However, Farrell and Shapiro are able to obtain general results about the “external” effect of the merger, that is, the combined effect of the merger on consumers and rivals. If we are prepared to presume that a proposed merger raises the combined profits of the merging firms (for otherwise they would not choose to merge), then any merger that generates positive external effects must raise total welfare. For a range of demand and cost conditions, Farrell and Shapiro provide an upper bound on the combined share of the merging firms such that their merger must generate positive external effects. If the combined share of the merging firms is small, they will not find it profitable to restrict output much, if at all, and when they do restrict output the larger firms are likely to expand, and shifting output towards larger firms actually boosts welfare, since the larger firms have lower pre-merger marginal costs. This approach has the significant virtue that it does not involve an inquiry into the efficiencies generated by the merger, which can be difficult to quantify and verify, as we discuss below.

Until now, we have examined the effects of mergers on price and welfare but have not related this analysis to the effect of the merger on industry concentration, a typical focus of horizontal merger enforcement policy (as reflected in the Merger Guidelines). Specifically, concern is typically thought to be greater, the higher is pre-merger concentration and the greater is the merger-induced increase in concentration, notably, as measured by the HHI. Farrell and Shapiro show, however, that increases in the HHI may well increase total welfare. In particular, they show that, starting from a Cournot equilibrium, an arbitrary small change in the outputs of all of the firms raises welfare if and only if $\frac{dX}{X} + \frac{1}{2} \frac{dH}{H} > 0$, where X , as before, is industry output.

Naturally, an increase in output raises welfare, since price is above marginal cost for all of the firms. More surprisingly, for a given change in total output, welfare is higher the greater is the change in concentration. Why? Each firm’s price-cost margin is proportional to its market share, so the larger firms have higher margins and thus lower marginal costs. As a result, shifting output towards them, which raises concentration, raises welfare as well. This observation tells us that an increase in concentration cannot serve as a proxy for a decrease in total welfare when studying horizontal mergers.⁹⁸ (It should be noted in this regard that, ordinarily, when enforcement agencies and courts consider increases in concentration, this is viewed diagnostically and prospectively under the maintained assumption that the share of the merged firm will equal the combined pre-merger shares of the merging firms.)

The applicability of the Cournot model is limited to industries where competition is accurately modeled as a quantity-setting game, or perhaps as a capacity-setting game followed by pricing competition, with fairly homogeneous products, and where the predictions of the one-shot Cournot model (rather than a model of repeated Cournot) fit the industry reasonably well. The Cournot model is not suitable for industries with highly differentiated products, especially if

⁹⁸ Farrell and Shapiro (1990a) show more generally how changes in the ownership of assets in Cournot oligopoly affects output, welfare, and the HHI.

capacity constraints are unimportant in the medium to long run. In those industries, a Bertrand model with differentiated products fits better. We now study mergers in that model.

2. Bertrand Model with Differentiated Products

A very extensive literature has developed to explore the effects of horizontal mergers in models of Bertrand competition with differentiated products.⁹⁹ These models are extensively used in practice to estimate and simulate the effects of proposed mergers, particularly in markets with branded products, ranging from consumer goods such as breakfast cereal to computer software.

Deneckere and Davidson (1985) provide a nice entrée point into this literature. In contrast to the results of Salant et al., they find that mergers are always profitable and will always involve price increases. Prior to the merger, the price of each product was set to maximize the profits earned on that product, given the prices of all other products. Now consider what happens if the price of one of the merging products, say product 1, is raised slightly. This will lower the profits earned on product 1, but the first-order effect will be zero since the price of product 1 was already optimized. The higher price for product 1 will, however, increase sales of product 2, thus raising the profits of the merged firm (a positive externality that firm 1 ignored prior to the merger). The increase in profits from product 2 will be larger the greater is the increase in sales of product 2 that results from the increase in the price of product 1 and the larger is the price-cost margin on product 2. What about changes in the prices set by the other firms? In Bertrand equilibrium, best-response curves slope upwards, so the other firms will find it optimal to raise their prices in response to the higher price for product 1 (and for product 2, the price of which it will also be profitable to increase). These higher prices increase the demand for products 1 and 2, further adding to the profits of the merged firm, which prospectively makes the merger even more attractive.¹⁰⁰ Note also that each non-merging firm welcomes the merger since it earns higher profits because, as explained, the merged firm charges higher prices for both of its products, which increases the demand for the rival products.

These ideas are very general: in models with differentiated products and Bertrand competition, mergers that involve no synergies are profitable for the merging firms, raise the prices charged by the merging firms, and raise the price and profits of the non-merging firms as well. Clearly, such mergers lower consumer surplus; they also tend to lower welfare. It is possible that such mergers raise welfare, however, if they involve significant synergies or if the merging firms are inefficient, so shifting output away from them and toward the other firms is efficient.

To apply these ideas in practice, where the emphasis tends to be on whether, and how much, a proposed merger will raise price, it is helpful to understand what economic variables tend to make the price effects of a merger between two suppliers of differentiated products large or

⁹⁹ See Ivaldi et al. (2003b), Motta (2004, pp. 243-265), and especially Werden and Froeb (2007) for more extensive reviews of this literature. Baker and Bresnahan (1985) is an important early contribution.

¹⁰⁰ The logic in the Cournot case is different because best-response functions slope down in that case. When the merged firm optimally reduces its output, the other firms expand output, which reduces the profits of the merged firm.

small. We return to this issue below, where we discuss the sophisticated simulation methods now used to estimate the price effects of such mergers.

A good sense of the basic forces at work can be gleaned by comparing the prices in a Bertrand duopoly with two differentiated products, each sold by one firm, with the price charged by a single firm selling both products. Focusing on just two products is not as restrictive as it might appear: one can interpret the demand functions for the two products in this model as demand in a general oligopolistic market, taking as given the prices of the all of the other products. In the absence of any efficiencies, the logic of Deneckere and Davidson (1985) tells us that the merged firm will have an incentive to raise its price, given the prices of the other firms, and that the optimal price for the merged firm, given those other prices, is less than the new Bertrand equilibrium price once one accounts for the price increases by the other firms. Therefore, the price increases calculated using a duopoly model will (somewhat) underestimate the price increases in the full oligopoly model.

We derived a formula in subsection II.C.2 for the difference between the monopoly price and the Bertrand equilibrium price in a simple, symmetric Bertrand duopoly model with linear demand and constant marginal cost. Following Shapiro (1996), we showed that the percentage gap

between the monopoly price and the Bertrand price is given by $\frac{P_M - P_B}{P_B} = \frac{\alpha}{2(1-\alpha)} \frac{P_B - MC}{P_B}$,

where $\alpha \equiv \frac{dX_2}{dP_1} / \left| \frac{dX_1}{dP_1} \right|$ is the diversion ratio, that is, the fraction of the lost unit sales of product 1, when the price of product 1 is raised, that are captured as unit sales of product 2, as previously defined in Section II.C.2. If we define the pre-merger price-cost margin as $\bar{m} \equiv \frac{P_B - MC}{P_B}$, then,

in this very simple model, the percentage price increase predicted from the merger of the two firms is $\frac{\alpha}{2(1-\alpha)} \bar{m}$.

This price increase is proportional to the pre-merger price-cost margin, \bar{m} . This comports with intuition: since the profits gained on the sale of product 2, when the price of product 1 is raised, are proportional to the margin on product 2, the magnitude of the margin on product 2 is proportional to the incentive to increase the price of product 1 (and conversely). Therefore, *ceteris paribus*, mergers between firms selling differentiated products are likely to raise price more, the greater are the pre-merger margins on their products.

The price increase associated with the merger is also proportional to the factor $\frac{\alpha}{1-\alpha}$, which is increasing in the diversion ratio and which is zero if the diversion ratio is zero.¹⁰¹ This, too, is intuitive: the greater is the diversion ratio, the greater is the share of the lost sales from product 1 that are captured by product 2 and thus internalized after the merger. Therefore, *ceteris paribus*,

¹⁰¹ We require $\alpha < 1$ or else the merged entity faces perfectly inelastic demand at all positive prices.

mergers between firms selling differentiated products are likely to raise price more, the closer is the degree of substitution between their products, as measured using the diversion ratio. Note that a high gross margin is consistent with a high diversion ratio; this pattern arises if the demand for product 1 is not very elastic and if a significant fraction of the (relatively few) sales lost when the price of product 1 rises are diverted to product 2.

As shown by Shapiro (1996), however, a rather different formula applies with constant-elasticity (rather than linear) demand. In this case, the percentage price increase predicted from the merger of the firms 1 and 2 is $\frac{\alpha \bar{m}}{1 - \alpha - \bar{m}}$.¹⁰² This ratio is larger than in the case of linear demand, and possibly much larger for plausible parameter values. To illustrate, suppose that the pre-merger gross margin is $\bar{m} = 0.35$, not an uncommon number for branded products, and that one-quarter of the sales lost when the price of product 1 is raised are captured by product 2 (and vice versa), so $\alpha = 0.25$. With these parameters, the post-merger price increase with linear demand is about 6%, while the post-merger price increase for constant-elasticity demand is nearly 22%.

Suppose that one observed the pre-merger margin of 35% and was able to estimate the diversion ratio of 25% between these two products. Both of these models—one with linear demand, one with constant elasticity of demand—can be parameterized to be consistent with these observations. Yet the two models give significantly different predictions for the price increase associated with a merger because the two demand systems diverge somewhat as prices depart from their pre-merger equilibrium levels. This should not be totally surprising: mergers are discrete events, and if nontrivial price changes are possible, their magnitude must in fact depend upon demand at prices distinctly different from the pre-merger prices.

All of this tells us that, in a merger involving differentiated products, making reliable predictions of unilateral price effects based on a model of Bertrand oligopoly requires an accurate structural model of the demand system, and that the shape of the demand system at prices some distance away from the pre-merger equilibrium affects the post-merger price increase.¹⁰³ If a structural model can be estimated that fits industry demand, then, with luck, the post-merger equilibrium can be simulated using that model, thereby predicting the magnitude of post-merger price increases. This promising approach has been explored analytically and applied in practice in recent years, as we discuss in subsection F.2. The logit model for differentiated products, championed by Werden and Froeb (2007), is especially tractable and has been used extensively to estimate the effects of horizontal mergers. In this model, each consumer picks one unit of a single brand from a set of choices that includes the differentiated products, $i = 1, \dots, N$, along with the alternative of an outside good, which can simply be interpreted as picking none of these products. The consumer's utility from selecting brand $i = 1, \dots, N$ is the sum of a "systematic" component associated with that brand, V_i , and an unobservable idiosyncratic component. Under suitable assumptions about the distribution of the idiosyncratic terms, the probability that a

¹⁰² We require $1 - \alpha - \bar{m} > 0$ so that the elasticity of demand facing the merged firm is greater than unity.

¹⁰³ All of these ideas carry over to mergers between multi-product firms, but the pertinent calculations are more complex.

consumer will pick brand i is given by $\phi_i = e^{V_i} / \sum_{k=0}^N e^{V_k}$, where the index zero corresponds to the outside good. If we define $\Phi = \sum_{k=1}^N \phi_k$ as the probability that the consumer will pick one of the N brands (rather than the outside good), then firm i 's market share is $S_i = \phi_i / \Phi$, so the market shares are proportional to the choice probabilities ϕ_i .

In the simple specification described in Werden and Froeb (2007), the systematic component of utility for brand $i = 1, \dots, N$ is given by $V_i = \gamma_i - \beta P_i$, where γ_i reflects the underlying quality or average attractiveness of brand i , P_i is the price of brand i , and β is a constant that determines the degree of substitutability among the different products. For large values of β , the competing brands are very close substitutes, and price-cost margins are low. Differentiating the demand for brand i with respect to the price of brand i gives $d\phi_i / dP_i = -\beta\phi_i(1 - \phi_i)$. Transforming this expression into elasticity form, the own-price elasticity for brand i is given by $-\beta P_i(1 - \phi_i)$. Differentiating the demand for brand i with respect to the price of brand j gives $d\phi_i / dP_j = \beta\phi_i\phi_j$. Transforming this expression into elasticity form, the cross-price elasticity of demand for brand i with respect to the price of brand j is $\beta P_j\phi_j$. Therefore, the diversion ratio from brand j to brand i , when the price of brand j rises, is given by $\frac{dX_i / dP_j}{|dX_j / dP_j|} = \frac{\phi_i}{1 - \phi_j}$.

This model has the attractive, but restrictive, property that the diversion ratio from brand j to brand i is proportional to firm i 's market share. In this important sense, the logit model is the antithesis of spatial models in which some products are very close substitutes, others are distant substitutes, and proximity need not bear any particular relationship to popularity. The logit model is a good starting point in a situation where all of the brands compete against one another and it is not clear which are "close" to each other. Nested logit models can be used when additional information about proximity is available.

Werden and Froeb (2007) show that, in the Bertrand equilibrium with single-product firms, the gap between firm i 's price and firm i 's marginal cost is given by $P_i - MC_i = \frac{1}{\beta(1 - \phi_i)}$. This

expression tells us that the firms with more attractive products, and thus larger market shares, have higher markups, much like firms with lower costs and thus larger shares have higher margins in a Cournot equilibrium. In the Bertrand equilibrium that results after the merger of brands 1 and 2, the equilibrium gap between price and marginal cost for each of the merging brands is given by $P_1 - MC_1 = P_2 - MC_2 = \frac{1}{\beta(1 - \phi_1 - \phi_2)}$. (Remember, in interpreting this

equation, that the market shares of the merging brands are not constants; they will fall as a result of the merger.) To illustrate, consider the situation in which product 2, say, is inherently more attractive, that is, in which $\gamma_2 > \gamma_1$. For simplicity, suppose that both products are produced at constant and equal marginal cost. Prior to the merger, product 2 would have a larger market

share and a larger gap between price and marginal cost than would product 1. After the merger, the prices of both products would be higher than their pre-merger levels, and the gap between price and marginal cost for the two products will be equal, which in this case further implies that the prices will be equal; product 2 would have a larger market share than product 1. Therefore, the post-merger price increase will be larger for product 1 than for product 2. This fits with intuition: the incentive to raise the price of product 1 is greater since a relatively large fraction of its sales will be diverted to product 2, due to that product's popularity. Furthermore, the pre-merger gap between price and marginal cost for product 2 is larger than that for product 1, so any diverted sales are actually adding to the profits of the merged entity.

The symmetric logit model with constant marginal cost can readily generate predictions about the price effects of mergers, given an estimate of the elasticity of demand for the market as a whole and an estimate of the pre-merger gaps between prices and cost. As an example, Werden and Froeb (2007) report that with six (symmetric) firms, a market elasticity of demand of -0.5 , a normalized pre-merger price of \$1, and a pre-merger gap between price and marginal cost of \$0.40, so marginal cost is \$0.60 (all this corresponding to $\beta \approx 2.9$), the merger of any two brands causes their prices to increase by about 6%. As they note, the logit model, with its lack of localization in competition, shows that a merger between two brands can easily raise price significantly even if the merging brands are not each other's next closest substitutes in any market-wide sense. Prices rise because, with only six firms, there are a nontrivial fraction of consumers for whom the merging brands are the first and second choices. In this model, the merged firm cannot identify and price discriminate against those consumers, so the merged firm raises price somewhat to all consumers.¹⁰⁴

Until now, the analysis has focused on the price effects of mergers that involve no production synergies. The consideration of efficiencies is facilitated by a convenient feature of models of Bertrand competition with differentiated products: the magnitude of the efficiencies necessary for a merger to reduce rather than raise price depends only upon the shape of the demand system (the diversion ratio between the merging products) at prices in the immediate neighborhood of the pre-merger equilibrium prices. A reduction in marginal cost of product 1 at the merged firm increases the gap between the firm's price and marginal cost on that product, giving the firm an incentive to lower its price. Price will in fact fall if this incentive is stronger than the incentive to raise the price of product 1 based on internalizing the diversion to product 2, now owned by the same firm. The direction of the price effect of the merger depends upon which of these effects is stronger. Since both of these effects are evaluated at the pre-merger prices, no information is required about the shape of the demand system at other prices.

Based on this logic, Werden (1996) derives an expression for the cost reductions necessary to prevent a merger from raising price.¹⁰⁵ In the symmetric case, where the two merging firms have

¹⁰⁴ If price discrimination were possible, the merged firm would raise price much more to the identifiable customers who ranked product 1 and product 2 as their first and second choice, and not at all to other customers. Effectively, one can compute a new post-merger equilibrium for each identifiable customer or group.

¹⁰⁵ This is the analogue in a Bertrand model of the necessary and sufficient condition for a merger to reduce price in Cournot oligopoly derived by Farrell and Shapiro (1990b).

equal market shares and gross margins prior to the merger, he shows that a merger will reduce price if and only if the cost reduction satisfies

$$\frac{MC_1 - MC_{12}}{MC_1} > \frac{\bar{m}}{1 - \bar{m}} \frac{\alpha}{1 - \alpha} \quad (7)$$

where $\bar{m} \equiv \frac{\bar{P}_1 - \overline{MC}_1}{\bar{P}_1}$ is again the pre-merger margin. This is a rather stringent condition in mergers between close rivals. Using our previous numerical example of $\bar{m} = 0.35$ and $\alpha = 0.25$, the merger must reduce marginal cost by about 18% to lead to a price reduction rather than a price increase. Note that reductions in fixed cost have no bearing on (short-run) price effects.

3. Bidding Models

In the Bertrand model, each firm sets a price, and buyers make their purchasing decisions given these prices. Bertrand models are especially well-suited for markets with differentiated consumer products in which there are a large number of relatively passive consumers.¹⁰⁶ In many other settings, however, there are large buyers who behave strategically, designing their procurement procedures so they can obtain the best price from their suppliers. In these settings, competition typically takes the form of bidding to win the business of a single customer who has designed a procurement procedure.

Many purchasing situations fit this pattern, including procurement auctions. The precise manner in which competition takes place depends upon the auction rules established by the customer. Klemperer (2004) provides an excellent overview of the enormous literature on auctions. Werden and Froeb (2007) discuss merger analysis in a situation where a seller is auctioning off an item using an ascending oral auction and the bidders have private values for the item. (Precisely the same ideas would arise in a situation where a buyer is running a procurement auction and the bidders are suppliers who differ in their costs of serving the buyer.) This auction format is equivalent to a second-price sealed-bid auction; it is a dominant strategy for each bidder to bid up to its value, and the price ultimately paid, P , is equal to the second-highest valuation among the bidders.

In this context, consider a merger between bidder 1 and bidder 2, and label the two bidders so that bidder 1's valuation, B_1 , is at least as large as bidder 2's valuation, B_2 . Label bidder 3 as the one with the highest valuation, B_3 , among the other bidders. A merger between bidder 1 and bidder 2 (which is equivalent here to collusion between these two bidders) will have no effect on the price paid for the item unless $B_2 > B_3$, that is, unless the two merging bidders have the two highest valuations on the item. If they do, price will fall from B_2 to B_3 . Viewed statistically,

¹⁰⁶ Even in markets for branded consumer products, large buyers such as large retailers may play a significant role. These buyers may be more active and strategic in dealing with manufacturers, in part by setting up bidding contests among their would-be suppliers.

merger effects depend on the joint distribution of the valuations of the bidders, including the merging bidders. Waehrer and Perry (2003) show how the price effect of a merger can be estimated for certain cumulative distributions of valuations.

B. Oligopoly Theory and Coordinated Effects

Mergers also can pose a risk to competition by increasing the likelihood that a collusive outcome will prevail. Such coordinated-effects theories of harm from horizontal mergers are featured in the Merger Guidelines, which state in §2.1: “A merger may diminish competition by enabling the firms selling in the relevant market more likely, more successfully, or more completely to engage in coordinated interaction that harms consumers.” Merger enforcement based on coordinated effects is more important, the more one believes that increased concentration contributes to coordinated outcomes and the less one believes that collusive behavior is readily deterred by antitrust law.

As discussed in Section III, collusion is generally thought to be easier to achieve and sustain when there are fewer suppliers in the industry. Therefore, at the simplest level, reducing the number of competitors by one tends to increase the likelihood of collusion. This idea underlies what is referred to as the *structural presumption*—that increases in concentration lead to less competitive interactions—that has long played a central role in antitrust. The heyday of the structural presumption corresponded with a time when industrial organization economists devoted substantial efforts to validating empirically the core idea of the structure-conduct-performance paradigm: markets that are highly concentrated tend to have higher prices and higher profits, and thus tend to serve consumers less well, than do markets with more competitive structures, *ceteris paribus*.

Demsetz (1973) mounted a strong attack on those who claimed that a positive cross-sectional relationship between concentration and profits was indicative of market failure or the need for an interventionist antitrust policy. Demsetz pointed out that a positive correlation would also arise if some firms were more efficient than their rivals, and if the more efficient firms had large market shares. Market concentration would then result from the presence of large, efficient firms. Under this hypothesis, small firms in concentrated market would earn normal profits, with the large, efficient firm earning profits due to Ricardian rents. If margins are associated with a firm’s market share, not overall market concentration, this may well reflect the greater efficiency of larger firms, at least in the short run. The implications for merger policy are profound: if a large firm seeks to buy a smaller rival, the resulting increase in concentration might go along with lower prices and consumer benefits if the large and efficient firm is able to improve the efficiency with which the assets of the smaller acquired firm are used. Bork (1978) is also well known for attacking the presumptions under merger policies of the 1960s and 1970s.

Reviewing the enormous literature on the cross-sectional industry relationships among concentration, prices, margins, and profits in order to distinguish among these competing hypotheses is beyond the scope of this Chapter. Schmalensee (1989) and Salinger (1990) are good starting places for readers interested in learning more. Pautler (2003) provides a more recent summary of the literature, which has made progress in distinguishing effects on a firm’s profits that are related to market concentration, the firm’s market share, or the firm’s identity (looking across multiple markets). Overall, economists have grown less confident over the past

several decades in stating that there is a systematic relationship between market concentration and market performance, at least over the range of market structures in which there are more than two or three firms. Even so, the cautionary statement made by Salinger in 1990 (p. 287) bears repeating today:

First, despite the well-known problems with this literature, it continues to affect antitrust policy. The inappropriate inferences used to justify an active antitrust policy have given way to equally incorrect inferences that have been used to justify a relaxed merger policy. Second, the alternative to cross-industry studies is to study specific industries. . . . [I]t is important to realize that it was the failure of studies of individual industries to yield general insights that made cross-industry studies popular.

Whatever one thinks of this literature, one should bear in mind that these cross-industry studies do not directly measure the effects of horizontal mergers, which we take up in Section C. The primary variation studied is across industries, not within an industry over time. Furthermore, through the early 1980s, highly concentrating horizontal mergers would simply not have been allowed. So, to the extent that one sees efficient larger firms in certain industries, through at least the early 1980s these firms largely arose through internal growth, non-horizontal mergers, or horizontal mergers involving relatively small market shares, not through highly concentrating horizontal mergers.

The key question regarding coordinated effects in merger analysis is whether a given merger will significantly increase the likelihood that a collusive outcome will arise.¹⁰⁷ In Section III, we explored in considerable detail how various industry conditions affect the likelihood of effective collusion. All of that theory and evidence can be brought to bear when considering coordinated effects in horizontal mergers. While we lack methods, such as we just discussed regarding unilateral effects, to quantify these effects, we know quite a lot about how a change in market structure resulting from a merger will affect the likelihood of effective collusion. In principle, then, one can trade off the increased costs from potential collusion against any efficiencies associated with a merger.

Some highly relevant and robust lessons emerged from the analysis of collusion in Section III. A horizontal merger between two significant suppliers, by reducing the number of players by one, can significantly increase the likelihood that the remaining firms will be able to reach a collusive agreement. One possibility is that a merger may establish a clear market leader who can play the role of price leader, serving the function of establishing and adjusting collusive prices, with the other firms following. Perhaps most important, when a firm that would have been reluctant to join in a collusive scheme (which, as previously noted, is sometimes termed a maverick) is acquired by another supplier who is larger or otherwise more inclined to participate, collusion can be greatly facilitated. Beyond these points, a merger reduces the number of bilateral links between firms in a market, which is some measure of the difficulty of reaching an agreement.

¹⁰⁷ Concern would also arise if the merger makes collusion more effective, for example, by raising the price at which collusion can be maintained to a level closer to the monopoly price or by reducing the frequency and duration of price wars. For simplicity, in our discussion below we use the shorthand of talking about the likelihood that a collusive outcome will arise.

With N suppliers, the number of such links is $N(N-1)/2$. A 5 to 4 merger reduces the number of links from 10 to 6; a 4 to 3 merger reduces the number of links from 6 down to 3.

For similar reasons, horizontal mergers also can make it easier to sustain a collusive outcome. A firm with a larger market share tends to have less to gain from cheating and more to lose if a price war erupts than do smaller firms. As a result, the merger of two smaller firms may increase the price at which collusion can be sustained. In general, a merger that significantly increases concentration will tend to make cheating on the collusive price less attractive, at least for the merging parties.

These observations are surely important for merger enforcement policy, even if our knowledge about the relationship between industry conditions and the likelihood of collusion does not give us a specific quantitative procedure to weigh the increased danger of collusion in, say, a 4 to 3 merger against efficiencies promised by that merger. However, the “paradox of proof” that we noted in subsection III.D.2 can present some problems when one seeks to apply collusion theory to horizontal mergers. To illustrate with an overly sharp example, suppose that one concludes in a given industry that effective collusion is quite unlikely if there are five or more firms, possible but not likely if there are four firms, and quite likely if there are three or fewer firms. Concerns about coordinated effects would therefore be minimal for any merger that left at least five firms in the industry. A merger from 5 to 4 firms would be a cause for concern, as would be a merger from 4 to 3 firms. But even more concentrating mergers, from 3 to 2 firms, and perhaps even a merger to monopoly, would cause fewer concerns: collusion is hypothesized to be likely with or without these mergers. While this is surely too strong a conclusion—even with only two firms, there probably is a nontrivial chance that collusion will break down—this logic at least undermines the standard presumption that mergers become more worrisome as the number of firms declines. This additional paradox is avoided only if one believes that probability of successful collusion is not just declining in the number of firms but also is a convex function of the number of firms.

There is relatively little formal theory exploring the implications for merger policy of the relationship between collusion and market concentration, apart from the papers already discussed in Section III. But several of the them are especially pertinent for evaluating coordinated effects in horizontal mergers. Notably, Compte et al. (2002) and Vasconcelos (2005) ask how the distribution of capacities affects the ability of the firms to sustain collusion in price-setting and quantity-setting supergames, respectively. And Davidson and Deneckere (1984) point out that reverting to the static Nash equilibrium typically is a less severe punishment when there are fewer firms (in a quantity-setting supergame or a price-setting supergame with capacity constraints), making for a complex relationship between market concentration and the likelihood of collusion.

Kovacic et al. (2006) propose an interesting new way to quantify the dangers associated with coordinated effects in a situation where a number of suppliers are bidding for the customer’s patronage. They propose measuring the effects of incremental collusion, that is, collusion that only involves two firms, before and after the proposed merger. They show how this calculation can be performed in a particular bidding model. While a large number of calculations are necessary to implement their method, these calculations are all well rooted in oligopoly theory, and in fact use the results already discussed in the analysis of unilateral effects.

Baker (2002) has emphasized the important role of maverick firms in disrupting or preventing collusion and thus the particular dangers that arise when a merger eliminates such a firm (an idea embraced in the Merger Guidelines as well). Collusion theory indicates that reaching an agreement and sustaining an agreement may be difficult if one of the firms expects to gain significant market share in the absence of collusion. Therefore, firms with strategies, products, or costs that are distinct from those of their rivals, and firms that are optimistic and growing rapidly, perhaps because they recently entered the market, are obvious candidates to be mavericks. Accordingly, Baker advocates an approach to merger enforcement policy that goes beyond the measurement of increases in market concentration by emphasizing the identification of mavericks. He argues that placing the focus on identifying maverick firms will reduce judicial errors by allowing the enforcement agencies and the courts to identify more accurately those mergers that are likely to have coordinated anticompetitive effects for any given level and change in market concentration. Plus, a merger may actually create a new maverick.

C. Empirical Evidence on the Effects of Horizontal Mergers

Given the large number of mergers that are consummated every year, including many horizontal mergers, one might think that there would be extensive, definitive evidence regarding the effects of these mergers. Under what circumstances have horizontal mergers been found to raise or lower prices, or more generally to benefit or harm consumers? And what has their impact been on the profits of the merging parties and on the profits of their rivals?

Sadly, there is no such clear and definitive body of evidence. To some extent, this reflects a lack of data: even in those relatively uncommon cases where one can accurately measure the prices charged before and after a merger, it may be hard to attribute price changes to the merger rather than to other changes in industry conditions. Also, the effects of a merger may arise in non-price dimensions such as product quality, customer service, or innovation. Furthermore, if merger enforcement policy is working well, the mergers most likely to have large adverse price effects are never proposed or are blocked on antitrust grounds. We do not mean to suggest that it is impossible to identify the effects of horizontal mergers; but nor is it easy. See FTC (2002) for some recent evidence.

Pautler (2003) offers an extensive review of empirical work on the effects of mergers and acquisitions. Readers interested in exploring this literature in greater detail should turn to his paper, which contains a treasure trove of information on the subject. Whinston (2006, pp. 110-127) also provides a valuable discussion of the evidence. We examine here several distinct methods for identifying and measuring the effects of mergers. In evaluating this evidence, one should bear in mind that over the past 25 years, only about 2% to 4% of the mergers reported every year under the Hart-Scott-Rodino Act were considered to raise sufficient antitrust issues to warrant a second request from the FTC or the DOJ, so data on the effects of all mergers may not

reflect the effects of the major horizontal mergers that are most likely to be scrutinized by antitrust authorities.¹⁰⁸

1. Stock Market Prices

One way to measure the effects of mergers is to study the stock market performance of the merging firms. Usually, this is done using an event study around the time of the announcement of the merger. This approach has been extensively explored in the finance literature. Andrade et al. (2001) provide an excellent introduction. The advantage of this approach is that it relies on detailed and accurate stock market data. However, by its nature, this approach cannot distinguish between favorable stock market returns based on efficiencies versus market power.¹⁰⁹ In addition, this approach measures the expectations of investors about merger effects, not the actual effects of mergers. Furthermore, this literature is not focused on horizontal mergers. Thus, the finance literature is best seen as addressing a more general question: do mergers and acquisitions produce wealth for shareholders or do they reflect managerial hubris? Finally, event studies do not readily disentangle predicted effects of the merger and other information that may be signaled by the announcement.

Andrade et al. (2001) report abnormal negative returns for acquiring firms, based on 1864 deals from the 1990s: 1.0% during a three-day windows around the announcement and 3.9% during a longer window from 20 days prior to the announcement through closing of the deal. However, target firms showed a 16% abnormal positive return during the three-day window. The combined firms gained about 1.5% over the short or longer window. They also report several studies that found negative abnormal returns over the three to five years following the completion of mergers, stating (p. 112): “In fact, some authors find that the long-term negative drift in acquiring firm stock prices overwhelms the positive combined stock price reaction at announcement, making the net wealth effect negative.” However, Andrade et al. are skeptical of these results, disputing the reliability of these longer-term studies, in part since it is hard to know what the “normal” return should be over these longer periods of time.

In the end, Andrade et. al. state (p. 117) : “We are inclined to defend the traditional view that mergers improve efficiency and that the gains to shareholders at merger announcement accurately reflect improved expectations of future cash flow performance. But the conclusion must be defended from several recent challenges.” One of these challenges arises from the fact that the source of the stock market gains to the combined firms from mergers has not been

¹⁰⁸ Pautler (2003) provides data on FTC and DOJ second requests. See FTC and DOJ (2003) and DOJ (2006) for more detailed recent data on merger challenges. Leary (2002) also provides some data on merger enforcement activities and merger activity.

¹⁰⁹ In principle, a merger that would lead to synergies and lower prices would depress the stock market value of rivals, while an anticompetitive merger that would lead to higher prices through unilateral or coordinated effects would boost the stock market value of rivals. Pautler (2003) reviews studies that attempt to measure the impact of horizontal mergers on the stock price of rivals. Such effects are more difficult to measure reliably than are effects on the stock market value of the merging parties, especially if the rivals are diversified companies with a relatively small share of their revenues coming from the sale of products in markets where the merging firms are significant horizontal rivals.

identified. In the case of horizontal mergers, at least, those gains could well come from enhanced market power. Another challenge arises because acquiring firms do not appear to benefit from mergers, which at the least is an uncomfortable fact for those who believe in a reasonably efficient stock market (unless one attributes many such mergers to managerial hubris or empire building). This will be important below when we consider merger synergies: if there truly are unique synergies resulting from the merger, why do acquiring firms fail to capture any of these gains from trade?

2. Accounting Measures of Firm Performance

A second method for measuring the effects of mergers is to study accounting data for the firms involved to look for changes in various measures, such as rates of return, cash flows, or profit margins. Ravenscraft and Scherer (1987, 1989), using widely cited FTC Line of Business Data, reach rather negative conclusions: many of the mergers and acquisitions they study were unsuccessful, leading to a decline in the post-merger profitability of the acquired line of business. Their study supports the view of excessive managerial zeal about acquisitions. However, they mostly examine conglomerate mergers, not horizontal mergers, so much of their evidence is not directly relevant to horizontal merger control policy. Also, they find that horizontal mergers tended to be more profitable than conglomerate mergers (although, again, this result does not distinguish market power from the possibility of greater synergies in horizontal mergers). Healy et al. (1992) examine post-merger operating performance for the fifty largest mergers that took place from 1979 to 1984. They find that the merged firms exhibited improved operating performance, as measured by operating cash flows, relative to their industry peers. They attribute these gains to increased operating efficiency. Along similar lines, Lichtenberg and Siegel (1987) and McGuckin and Nguyen (1995) find plant level productivity gains associated with mergers in manufacturing industries, using the Census Bureau's Longitudinal Establishment Data for 1972-1981. This was not a period, however, when highly concentrating horizontal mergers were permitted by antitrust enforcers.

3. Case Studies

A third approach is to study specific mergers, tracking the firms or industries involved, looking at such measures as prices, output, product quality, or R&D intensity. In principle, one can also try to measure the impact of a merger on rivals or customers. Kaplan (2000) provides a useful collection of case studies of mergers in a diverse set of industries, including hospitals, tires, banks, pharmaceutical drugs, airlines, and oil field services. The cases studied were not selected specifically to shed light on major horizontal mergers. These studies illustrate the great variety of fact patterns that arise in merger analysis, the important role of mergers as a means by which industry participants adjust to changing market conditions (making it especially hard to distinguish the effects of mergers from other changes taking place in the industry, especially once one recognizes that firms self-select to participate in mergers), and the risks as well as opportunities associated with mergers.

For antitrust purposes, it is most useful to study horizontal mergers that raised serious antitrust concerns when proposed but ultimately went forward. This approach has the virtue of focusing attention on the very small fraction of all mergers that are most relevant for assessing merger control policy.

Airline mergers have received a great deal of attention, in no small part because good data on fares are available and one can use fares on other routes as a good benchmark when measuring the effects of mergers on fares. Borenstein (1990), Werden, Joskow, and Johnson (1991), and Peters (2003) study two airline mergers from the mid-1980s that were approved by the Department of Transportation over the objections of the Department of Justice: the merger of Northwest Airlines with Republic Airlines, and the merger of Trans World Airlines with Ozark Airlines. These mergers raised significant antitrust issues because they combined directly competing hubs: Northwest and Republic both had hubs at Minneapolis, and TWA and Ozark both had hubs at St. Louis. Borenstein (1990) found significant fare increases following the Northwest/Republic merger but not following the TWA/Ozark merger. Werden, Joskow, and Johnson (1991) found that the Northwest/Republic merger raised fares by about 5% and the TWA/Ozark merger raised fares by about 1.5%, and that both mergers led to significant service reductions. Kim and Signal (1993) examine fourteen airline mergers from the mid-1980s. They compare price changes on the routes served by the merging firms with price changes on other routes of the same distance and conclude that any efficiency gains in mergers between rival airlines were more than offset by enhanced market power, leading to fares that averaged 10% higher after six to nine months. Fare increases were especially large for mergers involving airlines in bankruptcy, which had unusually low (perhaps unsustainably low) pre-merger fares.

The banking industry is another industry in which good price data are available and many horizontal mergers have occurred, making it possible to measure the price effects of horizontal mergers. Prager and Hannan (1998) study the effects of major horizontal mergers in the U.S. banking industry during the early 1990s. They look at changes in interest rates paid on deposits for several types of deposit accounts, using monthly data. They define “substantial horizontal mergers” as those that increase the HHI by more than 200 points to a post-merger value greater than 1800. They find that substantial horizontal mergers reduce the deposit interest rates offered by the merging banks.

Mergers have also been extensively studied in the hospital industry. The price and quality effects of hospital industry mergers have been examined in a number of studies, as described in Pautler (2003). For example, Vita and Sacher (2001) find large price increases, not reflecting increases in service quality, following a hospital merger in Santa Cruz, California.

Recent papers look at other industries as well. Pesendorfer (2003) studies the effect of horizontal mergers in the paper industry on capacity choices. Hastings (2004) looks at pricing in the retail gasoline market in Southern California.

One natural way to gain information to inform horizontal merger policy would be for the antitrust enforcement agencies to perform retrospective studies on the deals that they have investigated closely but ultimately allowed to proceed without significant divestitures. Neither the FTC nor the DOJ has officially reported results from any such study, at least in recent years.¹¹⁰ Barton and Sherman (1984) do report price increases from a highly concentrating

¹¹⁰ FTC (1999) reports on a study designed to determine the efficacy of the divestitures it had negotiated.

merger that was challenged by the FTC several years after it was consummated.¹¹¹ In addition, the U.K. Office of Fair Trading, in conjunction with the Department of Trade and Industry and the U.K. Competition Commission, sponsored a study of ten mergers that took place during 1990-2002. These were mergers that the Office of Fair Trading had reviewed and found to raise sufficient competition issues that they were worthy of referral to the Competition Commission but that the Competition Commission had subsequently cleared. See Office of Fair Trading (2003). Based on interviews with customers of the merging firms, this study did not find a significant lessening of competition in eight of the ten cases studied. In the other two cases, a short-term loss of competition was found to have been corrected by subsequent entry into the market.

D. Antitrust Law on Horizontal Mergers

This section outlines current antitrust law on horizontal mergers. We start with a brief statutory background and an explanation of procedures, emphasizing pre-merger notification and analysis by enforcement agencies. We then discuss the substantive law regarding requisite anticompetitive effects and whether merger synergies may be offered to defend otherwise anticompetitive mergers.

Throughout the discussion, it is useful to keep in mind the relationship between merger law, on one hand, and the law concerning price-fixing and monopolization, on the other hand. Because collusion is difficult to detect and prosecute (and, depending on the means of collusion, is of uncertain illegality), as discussed in Section III, it makes sense to some degree for merger policy to adopt a prophylactic approach toward mergers that threaten greater cooperation among firms. Likewise, because the law on monopolization does not regulate price-setting once a merger has been validated and imposes only modest constraints on exclusionary practices, as will be discussed in Section V, there is also reason to be wary of approving a merger that threatens unilateral effects or exclusionary conduct.

1. Background and Procedure

As elsewhere, our discussion will focus on antitrust law and procedures in the United States; there is a growing but incomplete convergence in how horizontal mergers are treated across jurisdictions.¹¹² Relevant U.S. law has three primary, overlapping provisions: Sherman Act Section 1's prohibition on any "contract, combination . . . , or conspiracy in restraint of trade" (the focus in Section III on collusion), Clayton Act Section 7's prohibition on acquisitions of stock or assets whose effect "may be substantially to lessen competition, or to tend to create a

¹¹¹ The acquiring company was Xidex, and the products involved were types of duplicating microfilm.

¹¹² In 2004, the European Union promulgated new horizontal merger guidelines that in many respects are similar to the preexisting Merger Guidelines in the United States. European Union (2004b). Prior to that, although few mergers had been blocked, the enforcement stance of the Commission is generally regarded to have been stricter than that in the United States. The European Court of First Instance's reversal in 2002 of three Commission attempts to block mergers is seen as the catalyst for the recent reform. Other notable administrative changes include the appointment of a chief competition economist. See, for example, Dabbah (2004). It is too early to tell just how much practice under the new regime will differ in fact from that in the United States

monopoly,” and the Federal Trade Commission Act Section 5’s prohibition on any “unfair method of competition.”¹¹³ In spite of diverse histories and statutory language, a largely unified approach to enforcement of these provisions has emerged. Notably, the DOJ and FTC in their most recent Horizontal Merger Guidelines (1992) have promulgated a single policy statement applicable regardless of the statute involved. Commentators and courts have largely taken a similar approach.¹¹⁴

Most challenges to mergers are brought by one of the two federal agencies, although mergers may also be challenged by states and private parties.¹¹⁵ Since 1976, the federal procedure has taken its current form, which is similar to procedures in many other jurisdictions.¹¹⁶ Firms intending to merge are required to file specified information with the pertinent agencies. Each deal is cleared to either the FTC or the DOJ. In mergers for which there is any serious prospect of a challenge, the parties usually submit substantial supplemental material. They hire a team of lawyers and economic experts (often associated with consulting firms) that typically have substantial experience in merger filings; indeed they may have handled numerous prior mergers in the same or related industries. This team gathers and analyzes information and produces an often-elaborate study document defending the merger with regard to competitive effects and anticipated efficiencies. The goal is to persuade the agencies to approve the merger, and to do so promptly.

An important aspect of the procedure concerns the effects of agency delay—which arises when the agency feels that it needs additional information or must undertake more substantial independent investigation and analysis—or of an ultimate agency challenge. Even if the parties anticipate eventual approval, whether from the agency or after litigation in court, the prospect of delay will kill many deals and impose substantial costs on others. Keeping financing in line, making interim investment decisions in plant and equipment, deciding on strategic matters such as launching new products or terminating old ones, maintaining customer loyalty in the presence of uncertainty about product support, retaining talented employees who may fear job loss, and so

¹¹³ Observe that none of the statutes is limited to mergers per se; other forms of combination, notably including acquisitions of some or all of another firm’s assets, are included. (Thus, for example, if the only two products in a market are patented, the acquisition of one of the patents by the owner of the other would be analyzed similarly to a horizontal merger.) EU regulations have a similar reach.

¹¹⁴ See, for example, Areeda, Hovenkamp, and Solow (1998, vol. 4, revised ed., pp. 42-43).

¹¹⁵ One might think that competitors would frequently challenge mergers. However, under the doctrine of “antitrust injury,” this is not ordinarily possible: competitors tend to be injured by pro-competitive mergers that lead to lower prices (deemed not the sort of injury that the antitrust laws were enacted to prevent) but helped by anticompetitive mergers (recall from Section A, for example, that unilateral effects tend to benefit non-merging parties). Relatedly, when agencies are investigating proposed mergers, they are less likely to give weight to the views of competitors on overall effects (for fear of manipulation) and more commonly seek the reactions of large purchasers (for example, health insurers, in the case of hospital mergers). In this regard, Coate and Ulrick (2005) find that the probability that the FTC takes action against mergers is higher, *ceteris paribus*, when there are customer complaints about the merger. In recent commentary, the agencies affirm that “Consumers typically are the best source, and in some cases they may be the only source, of critical information...” DOJ and FTC (2006, pp. 9-10).

¹¹⁶ Indeed, there has been some explicit international cooperation, motivated by the fact that many substantial mergers are subject to the competition regulation of multiple national and international jurisdictions.

forth may present significant challenges to merging firms, especially those being acquired, that do not know if or when their deal may be approved. Accordingly, great energy is devoted to obtaining a quick and successful conclusion to the antitrust agency's deliberations. If the agency challenges the deal and no settlement is reached, the merging parties must either abandon their transaction or confront considerable further delay in the process of litigating the matter in federal court.

To give some sense of the level of merger review activity, during the years 2001-2005 for the Department of Justice there were about 1000 to 2400 pre-merger notifications annually, of which 70 to 106 resulted in decisions to investigate further, and 2 to 7 led to cases being filed, depending on the year.¹¹⁷ As already suggested, however, these latter statistics can be misleading because some mergers will be dropped along the way either because the parties are insufficiently confident of success or simply because they cannot tolerate the anticipated delays. Also, no doubt, some potential mergers are deterred; the more predictable are the agencies, due in part to the Merger Guidelines and years of experience under them, the more one would expect there to be few proposed mergers with a high likelihood of being challenged.

Another important outcome is settlement, most frequently through the parties spinning off plants, other operations, or lines of business in areas of significant competitive overlap.¹¹⁸ That is, some mergers may be found to pose a serious competitive threat but only in certain geographic markets or only with respect to some of the many products the firms produce. In such cases, appropriate divestiture of pertinent assets will ordinarily satisfy the enforcement agencies.

For challenges that do proceed to court, the agencies often attempt to obtain a preliminary injunction requiring the merging firms to continue to operate independently pending a final outcome, and in the end successful challenges produce permanent injunctions against the merger (or subsequent negotiations leading to asset divestitures). This approach contrasts dramatically with the course of proceedings in earlier years, before the pre-merger notification regime was in place. Then, mergers were promptly consummated, and final decrees against challenged mergers ordinarily took many years, sometimes more than a decade, at which point the two firms were often sufficiently integrated (plants closed, brands discontinued, new joint operations well underway) to make practical divestiture difficult or impossible.

The remainder of this section focuses on the substance of the legal restriction on horizontal mergers, first examining the core inquiry into anticompetitive effects and then considering the role of efficiencies in justifying mergers that would otherwise be proscribed. As will be seen, the

¹¹⁷ DOJ (2006). The FTC's merger enforcement activity is comparable to that of the DOJ. In the European Union, from 1990 until May 2002 (and thus before the promulgation of the new 2004 regulations and guidelines), 86% of notified mergers were approved unconditionally, 5% were approved subject to undertakings (such as spin-offs) by the end of the one-month initial investigative period, an additional 3% after further investigation, and 1% (18) were prohibited. (Another 1% were withdrawn during in-depth investigations; various others were found to be outside the Commission's jurisdiction.) See Walle de Ghelcke and van Gerven (2004, § 5.01).

¹¹⁸ During the past decade, the combined number of transactions that were restructured or abandoned after a formal challenge was announced but before a case was filed in court usually exceeded the number of cases filed. And those statistics include only terminations that followed the issuance of a formal challenge.

approach toward both issues has evolved a great deal over time. Furthermore, throughout this evolution these two issues have not been entirely independent. In particular, the central question of the likelihood and extent of anticompetitive effects required to condemn horizontal mergers—the threshold for a successful challenge—seems to be answered in a manner that substantially reflects underlying views about the typical probability and magnitude of merger synergies.

2. Anticompetitive Effects

In the 1960s (in the wake of the strengthening of Clayton Act Section 7 in 1950), the U.S. Supreme Court, following the lead of the federal enforcement authorities, adopted a restrictive view toward horizontal mergers. The Court condemned a number of mergers where the parties' combined market shares were under 10%, for example, in *Brown Shoe* and *Von's*.¹¹⁹ The first government merger guidelines, promulgated in 1968, adopted similarly stringent thresholds for challenging mergers. Likewise, they endorsed the structural presumption that concentration implies anticompetitive effects, as articulated by the Supreme Court in *Philadelphia Bank*: “a merger which produces a firm controlling an undue percentage share of the relevant market, and results in a significant increase in the concentration of firms in that market is so inherently likely to lessen competition substantially that it must be enjoined in the absence of evidence clearly showing that the merger is not likely to have such anticompetitive effects.”¹²⁰ A shift was signaled by the 1974 decision in *General Dynamics*.¹²¹ The specific holding—that the *prima facie* case established by market share statistics could be rebutted by showing that the figures gave a misleading depiction of competitive effects—was not itself truly novel (*Brown Shoe* had suggested as much). However, the acts of subjecting the government's case to heightened scrutiny and ultimately rejecting it were taken as a signal of a new direction.

Since the mid-1970s, there have not been further merger opinions by the Supreme Court.¹²² Nevertheless, a confluence of three factors has made clear that the law has moved substantially: changing views toward competition and the effectiveness of market forces (both broadly and in the academy), a change in the composition of the Supreme Court and in the nature of its opinions

¹¹⁹ *Brown Shoe Co. v. United States*, 370 U.S. 294 (1962); *United States v. Von's Grocery Co.*, 384 U.S. 270 (1966). Indeed, in *Von's*, after pointing out that the combined share was 7.5% of the Los Angeles market for grocery stores, that the number of independent stores had fallen from 5365 to 3590, that half of the top 20 chains had acquired stores from smaller firms, and similar facts, the Court proclaimed: “These facts alone are enough to cause us to conclude . . . that the Von's-Shopping Bag merger did violate §7.” 384 U.S. at 274. The dissent criticized the majority for attempting to “roll back the supermarket revolution” and asserted that “[t]he sole consistency that I can find is that under §7, the Government always wins.” 384 U.S. at 288, 301.

¹²⁰ *United States v. Philadelphia National Bank*, 374 U.S. 321, 363 (1963). The Court cited prominent economic and legal authorities in support of this view, although the market share levels in *Philadelphia Bank* and proposed by most of the commentators were substantially higher (20% or more) than the levels deemed sufficient in many of the other cases of the period and in the 1968 guidelines.

¹²¹ *United States v. General Dynamics Corp.*, 415 U.S. 486 (1974).

¹²² This apparent anomaly is largely explained by two procedural changes: the new requirement of pre-merger notification reduced the flow of questionable mergers into the courts, and the elimination of special rules permitting automatic appeal in some cases directly to the Supreme Court greatly reduced the proportion of antitrust cases reaching that court.

on other antitrust subjects, and a new direction from the government as embodied in the 1982 merger guidelines. The current (1992) Merger Guidelines are a successor to the 1982 version, which differed in many key respects from those issued in 1968. Details of the current methodology for evaluating horizontal mergers will be examined throughout the remainder of this section.¹²³ Perhaps the most notable change, however, was in the thresholds for challenge: they were notably higher, sufficiently so that a number of the famous cases of the 1960s (that the government won) would not have been brought had the new guidelines been in effect.

The presumptive thresholds in the Merger Guidelines (once the market is defined; see Section E) are as follows. If the post-merger HHI is below 1000, the market is regarded as unconcentrated and ordinarily no further analysis will be undertaken. If the post-merger HHI is between 1000 and 1800, significant concerns are raised if and only if the merger increases the HHI by more than 100, in which case further analysis is undertaken.¹²⁴ And if the post-merger HHI exceeds 1800, significant concerns are deemed to exist when the merger raises the HHI by more than 50.¹²⁵ This air of certainty is misleading. In actual application of the Merger Guidelines, it has become apparent that, in certain industries, the *de facto* thresholds are much higher. For example, most of hundreds of hospital mergers subsequent to the 1982 guidelines (which had the same thresholds as those just described) have gone unchallenged even though post-merger HHIs and HHI increases were greatly above 1800 and 50, respectively, presumably reflecting a view of typical efficient scale in this industry.¹²⁶ Even so, the lawyers and economists who specialize in merger practice are generally aware of such patterns, which themselves probably reflect successful persuasion in prior merger filings in particular industries.¹²⁷

¹²³ For further elaboration on how the guidelines are implemented in practice, see the enforcement agencies' commentary, DOJ and FTC (2006).

¹²⁴ Further analysis involves consideration of entry, efficiencies, and the possibility that one of the firms may be failing.

¹²⁵ The new 2004 EU guidelines (European Union 2004a) are strikingly similar: 2000 replaces 1800, and the increases must be 250 and 150 rather than 100 and 50, respectively. As noted in the text and notes to follow, however, actual practice under the U.S. Merger Guidelines indicates the use of higher *de facto* thresholds (to an extent that varies by industry). Likewise, the additional factors considered in the new EU guidelines are largely the same as in the United States—notably, they include efficiencies as a defense—and they even appear in the same order.

¹²⁶ See also DOJ and FTC (2003) which gives statistics on the post-merger HHI and change in HHI for merger challenges in a number of industries. The tables strongly suggest that the thresholds for challenges vary greatly by industry. Coate and Ulrick's (2005) analysis of merger enforcement at the FTC finds that, for given HHIs and other factors, enforcement actions are more likely in the oil, grocery, and chemical industries. In addition, they find that, holding both the HHI and change in HHI constant, the probability of enforcement rises as the number of leading rivals falls from 5 to 4 and from 4 to 3.

¹²⁷ Leary (2002) emphasizes that, since the Merger Guidelines have now been followed for years, current merger policy exhibits a good deal of stability. Furthermore, in order "to provide greater transparency and foster deeper understanding regarding antitrust law enforcement," the DOJ and FTC recently issued a document, "Commentary on the Horizontal Merger Guidelines" in 2006. The agencies also held a Merger Enforcement Workshop in 2004; extensive materials from that workshop are available at www.usdoj.gov/atr/public/workshops/mewagenda2.htm.

Even though the Merger Guidelines are formally just a public statement by the federal agencies of how they intend to proceed internally, they have to a substantial degree dictated parties' practices in litigation and courts' analyses of mergers regarding the methodology for assessing market definition, anticompetitive effects, and other factors as well as the thresholds for condemnation. The courts undoubtedly welcome the guidance. Moreover, as noted, even though the most on-point Supreme Court opinions (from the 1960s) are much stricter, there has been a sufficient shift in understandings and in behavior of the Supreme Court on other issues that lower courts have followed the new approach rather than adhering to older precedents. The most direct indicator of this fact is that, in this more recent era, the government loses a good proportion of the cases it brings even though the government's threshold for challenging mergers is much stricter than in the past (when they won nearly every case).

3. Efficiencies

Merger synergies play an important role in competition policy toward horizontal mergers, and one that has changed substantially over time. In this subsection, we first describe that role and its modern development and then consider how it relates to the goals of competition policy more generally.

Although efficiencies are usually discussed as a possible defense, advanced by the merging parties to justify a merger that might otherwise be condemned as anticompetitive, efficiencies have long had another, more significant influence on merger policy: in setting the threshold for antitrust scrutiny. Consider trivial combinations, say, when two individuals form a partnership or two small stores join forces. These combinations as well as some substantially larger mergers have never been subject to challenge. But they could have been. First, as noted in the discussion in subsection III.E.2, such combinations literally involve price-fixing, going forward, which is automatically condemned, supposedly without proof of market power. Yet productive combinations, under the rule of reason, are permissible, presumably because they often create synergies (despite the fact that, once formed into a single entity, prices will be determined jointly rather than independently). Second, even when the number of firms is large, mergers in Cournot or Bertrand oligopoly tend to raise prices (slightly). Nevertheless, nontrivial anticompetitive effects must be demonstrated before a merger will be challenged.¹²⁸

Viewed more broadly, setting the threshold of anticompetitive effects significantly above zero may be rationalized by the view that mergers typically generate some synergies, so they should not be prohibited unless the reduction in competition is sufficiently great. In the 1960s and 1970s when U.S. anti-merger policy was strict, many fairly large mergers were nevertheless routinely permitted. Currently, as described in subsection 2, the thresholds are much higher; moreover, they are raised further in industries where synergies are thought to be unusually large relative to the size of the market (hospitals, for example).

¹²⁸ Although the language of the Clayton Act demands a substantial effect, that of the Sherman Act does not. (Historically, the Clayton Act, passed in 1914 to remedy perceived weakness in the Sherman Act and amended in 1950 to strengthen it further, has been viewed as stricter than the Sherman Act. As noted in subsection 1, however, the enforcement agencies and, increasingly, the courts apply a more unitary approach.)

Therefore, it seems appropriate to understand an efficiencies defense to a merger whose suspected anticompetitive effects exceed the threshold as implicitly involving a claim that the merger synergies are not merely substantial but are large enough to notably exceed the level ordinarily presumed to exist. After all, they must be enough to justify the merger in light of what would otherwise be substantial, not merely trivial, anticompetitive effects. This framing of the question may help explain why courts and enforcement agencies are cautious in accepting efficiency defenses (in addition to the obvious reason that merging parties have every incentive to assert the existence of synergies when there are few and the merger is in fact anticompetitive).

In the 1960s, the U.S. Supreme Court exhibited a somewhat schizophrenic but ultimately hostile attitude toward merger synergies. Most famously, in *Brown Shoe*, the Court viewed the efficiencies resulting from the vertical aspects of the merger as problematic because they would give the combined entity an advantage against competitors. On one hand, the Court stated: “It is competition, not competitors, which the Act protects.”¹²⁹ However, the passage continues: “But we cannot fail to recognize Congress’ desire to promote competition through the protection of viable, small, locally owned businesses. Congress appreciated that occasional higher costs and prices might result from the maintenance of fragmented industries and markets. It resolved these competing considerations in favor of decentralization. We must give effect to that decision.”

Over time, the first statement—“competition, not competitors”—has continued to be among the most-quoted passages from all Supreme Court antitrust opinions, by courts and commentators alike, whereas the latter clarification (contradiction) is not usually appended anymore. Lower courts in the last couple of decades (recall, there have been no recent Supreme Court pronouncements on mergers) have varied in their approaches, some expressing uncertainty about an efficiencies defense but most accepting it, at least in principle.¹³⁰

The Merger Guidelines are now unequivocal. In 1997, amendments to the 1992 version added a new section on efficiencies. They state: “Indeed, the primary benefit of mergers to the economy is their potential to generate such efficiencies. Efficiencies generated through merger can enhance the merged firm’s ability and incentive to compete, which may result in lower prices, improved quality, enhanced service, or new products.” Two main conditions are imposed: only merger-specific efficiencies are recognized, and those efficiencies must be sufficiently verifiable. Cognizable efficiencies also must be sufficiently large to prevent the merger from being anticompetitive.¹³¹

Which efficiencies are merger specific? Under the agencies standard approach, economies of scale from, say, building larger plants, are likely to be accepted. Efficiencies from some other functions, like combining payroll operations, are less likely to be credited because of the option of contracting out for such services, achieving benefits of scale short of merger. And if two hospitals demonstrate a need to share modern imaging equipment due to high fixed costs, they

¹²⁹ *Brown Shoe Co. v. United States*, 370 U.S. 294, 344 (1962).

¹³⁰ See, for example, the cases cited in Areeda, Hovenkamp, and Solow (1998, vol. 4A, p. 29 n.17).

¹³¹ As noted previously, the 2004 EU guidelines are quite similar, including a recognition of an efficiencies defense and the particulars of how it may be established in a given case.

may be permitted to form a joint venture limited to that purpose but not allowed on this ground to go forward with an otherwise anticompetitive merger of all of their operations.

Farrell and Shapiro (2001) explore which efficiencies should qualify as being merger specific, recognizing that in the presence of economies of scale firms can grow internally to reduce their average costs. They distinguish between efficiencies based solely on scale economies and efficiencies that reflect the combination of specific, non-tradable assets owned by the merging parties, which they consider true merger-specific efficiencies, or synergies. They argue that many claimed efficiencies are not in fact merger-specific.

There is a deeper problem underlying many disputes about whether efficiencies are merger specific. Often, the parties will claim that the merger is necessary whereas enforcers will question why the purported benefit cannot be achieved through some more limited form of contractual arrangement. Under what circumstances there exist benefits from combining activities under the direction of a single firm that cannot be achieved through contracting is, of course, a core issue in the theory of the firm, one whose exploration was launched by Coase (1937), extended by Williamson (1975, 1985), and explored in subsequent work by Grossman and Hart (1986), Holmstrom and Tirole (1989), Hart and Moore (1990), and Hart (1995), among many others. Since the findings of this literature are often subtle, depending on factors that may not be readily and reliably ascertained in the course of an antitrust dispute, there will be an inevitable area of uncertainty. The level of theoretical sophistication reflected in modern analysis of contracts and the firm has yet to make significant inroads in merger analysis. In practice, the agencies tend to look at the forms of collaboration short of merger that are actually used in the industry to determine whether certain claimed efficiencies are merger specific.

In the end, enforcers do and should have a healthy skepticism about self-serving efficiency claims made by competitors seeking to merge. As we noted in Section C, acquiring firms often seem to overestimate the benefits of an acquisition, and the overall record regarding merger efficiencies is mixed at best. And it tends to be very difficult for the enforcement agencies or the courts to assess whether claimed efficiencies are indeed likely to arise. Kolasky and Dick (2003) examine the treatment of efficiencies in merger cases, arguing that the enforcement agencies and the courts have made substantial progress over the past twenty years incorporating efficiencies into merger analysis.¹³²

Suppose that synergies can be quantified, or that in a particular case the enforcement agency or the court has done its best in determining the matter. The important remaining question is how efficiencies and anticompetitive effects are to be balanced. Should the decision depend on total economic welfare—the sum of consumer and producer surplus—the standard normative economic approach? Should it instead turn solely on consumer surplus—whether prices rise or fall—that is, whether the efficiencies are sufficiently passed on to consumers to at least offset any price increase that would otherwise ensue due to anticompetitive effects? Or should some other standard be used, which seems to be the approach in earlier cases like *Brown Shoe*?

¹³² See also Pitofsky (1999) and Muris (1999) for different views on the role of efficiencies in merger analysis.

This set of questions brings to mind the discussion in subsection III.E.2 on the meaning of the rule of reason.¹³³ Recall that the *Chicago Board* case and subsequent decisions define reasonableness in terms of what promotes versus suppresses competition. Yet there remained ambiguity concerning the meaning of competition.

Under a process-oriented view, which has support in many modern cases, one might be concerned with protecting rivalry *per se*, which might imply a strict merger policy. That view seems consonant with the 1960s merger decisions and language such as that quoted above from *Brown Shoe* (the latter part, referring to the merits of protecting competitors and preserving decentralization) or that in *Philadelphia Bank* referring to the preservation of a “traditionally competitive economy.”¹³⁴ Under this approach, not only would one condemn a merger from 2 firms to 1, 3 firms to 2, and 4 firms to 3, but also, it might seem, from N firms to $N - 1$, no matter how large was N . But as discussed previously, even when this view of merger policy was dominant, smaller mergers were permitted. Although efficiencies were not recognized as a defense and were sometimes viewed as an evil in larger mergers, the threshold for challenge was high enough to allow countless mergers to go unchallenged. Furthermore, modern cases such as *BMI* interpreting the rule of reason and the bulk of lower court cases on mergers, as well as the Merger Guidelines, accept more explicitly that efficiency counts positively, and in particular that better serving customers, notably, through lower prices, is desirable.¹³⁵

Thus, at present the main contest seems to be between consumer welfare and total welfare, that is, whether efficiencies should be credited when they increase producer surplus rather than being passed on to consumers. The influence of merger synergies on price depends on both the manner in which the firms’ cost function is altered by the merger and on the nature of firms’ interaction. At one extreme, if a merger produces no (or negligible) cost reductions but reduces the number of competitors in a Cournot or Bertrand oligopoly, and N is not very large, prices will rise nontrivially and the reduction in consumer welfare will be approximately equal to the fall in total welfare. No tradeoff needs to be considered. To take another possibility, suppose that two firms merge to monopoly and that all the savings are in fixed costs; then prices will rise (unless there was perfect collusion previously) because fixed costs do not ordinarily affect pricing decisions. However, if the increase in deadweight loss is not that great but the savings in fixed costs is large, then consumer welfare may fall while total welfare rises. For a further contrast, consider a merger of two firms in a setting with many firms and assume that the cost saving involves a reduction in marginal cost; then the merged firm may price sufficiently more aggressively to bring prices down, in which case both consumer surplus and total welfare would rise. Viewed

¹³³ The rule of reason, note, arose in interpretation of Sherman Act Section 1, which is also one of the statutes applicable to horizontal mergers. Thus, as a formal legal matter, one might say that the rule of reason is the Sherman Act standard for horizontal mergers. However, courts have generally considered mergers separately, and as noted, increasingly without regard to which statute is invoked in a particular case. Nevertheless, given the growing convergence in approaches under all of the antitrust statutes, one should expect some congruence between courts’ analyses of horizontal mergers and of what are denominated as rule of reason cases.

¹³⁴ *United States v. Philadelphia National Bank*, 374 U.S. 321, 371 (1963).

¹³⁵ Other jurisdictions, with different laws and histories, might give weight to other objectives in addition to or instead of economic welfare (whether consumer welfare or total welfare). For example, antitrust policy in the European Union has traditionally placed some weight on the value of integration into a single market.

more broadly, the analysis in Sections A and B above (the latter of which drew on Section III on collusion) elaborates conditions under which prices may be expected to rise or fall and how those conditions depend on the manner in which a merger may affect firms' costs.

Accordingly, although many mergers raise both consumer and total surplus or reduce them both, there will exist a notable subset of mergers that increase total surplus but reduce consumer surplus. The resulting need to choose between consumer welfare and total welfare as guides to merger policy was presented sharply by Williamson's (1968) classic discussion of the tradeoff between market power and efficiencies and, in particular, his demonstration in a basic case that even modest gains in productive efficiency could exceed the losses in allocative efficiency from price increases.¹³⁶ Recent contributions to the debate about the proper objective of antitrust policy include Farrell and Katz (2006), Heyer (2006), Kolasky and Dick (2003), and Salop (2005).

The precise language of the various statutes—which as prior discussions indicate are not taken literally or interpreted independently—gives conflicting guidance in determining the underlying principle.¹³⁷ Clayton Act Section 7's prohibition on mergers that may substantially lessen competition or create monopoly might be interpreted to prohibit any mergers that reduce rivalry or at least those leading to high concentration, and certainly to a single firm serving the market. Yet it is possible that even a merger to monopoly could raise both total welfare and consumer welfare. The Sherman Act's prohibition on restraints of trade is interpreted under the rubric of the rule of reason, which has been explained to contain substantial ambiguity even after being translated into the promote/suppress competition test and refined through modern applications. The FTC Act's prohibition of "unfair" competition is vague and question-begging on its face.

That said, the modern trend in the United States seems to be toward a consumer welfare standard when considering the efficiencies defense, although the legal authorities have not elaborately rationalized this view, specifically by defending it against the alternative of total welfare. The Merger Guidelines' discussion of efficiencies at one point refers to those that "likely would be sufficient to reverse the merger's potential to harm consumers in the relevant market, for example, by preventing price increases in that market." Likewise, a number of lower courts

¹³⁶ In a market where prices are initially at a competitive level, the benefits from productive efficiency are a rectangle (assuming a uniform downward shift in the marginal cost curve) whereas deadweight loss is the familiar triangle, which for small price increases will be a much smaller quantity. However, if price is nontrivially above marginal cost before a merger, as might be expected if pre-merger concentration is high, the incremental deadweight loss from price increases will be larger. Also, cost savings from mergers, when they exist, can take many different forms, so the benefit need not be indicated by a rectangle (that is, quantity times a uniform change in marginal cost).

¹³⁷ The legislative histories, however, are clearer. When the Sherman Act was enacted in 1890, it is unimaginable that the legislators appreciated modern concepts of efficiency and deadweight loss, which were unknown even to most economists at that time. Legislators did seem to care about high prices and also about the protection of small businesses. As discussed by the Supreme Court in *Brown Shoe*, the Clayton Act amendments in 1950 that led to the version of Section 7 that is close to its current form were motivated by general concerns (largely of a political and social sort) about concentration of power, most of which seem disconnected from either consumer or total economic welfare. As indicated by the text in this section and in Section III, however, these views have not significantly influenced courts' decisions in recent decades.

considering the issue have indicated the need for the merging firms to demonstrate that efficiencies be passed along to consumers.¹³⁸

From an economic point of view, however, it would seem that in principle total welfare should be the standard.¹³⁹ Producers have owners who are people, just like final consumers. One might nevertheless favor consumers on distributive grounds because owners are on average richer than consumers (although obviously the groups overlap substantially). However, it is usually more efficient to achieve distributive objectives directly, through the income tax and transfer system.¹⁴⁰ In addition, one should add that, in antitrust, the protected class is often producers. This is patently so when the merging parties sell intermediate goods (although savings to purchasing firms that operate a competitive market will ultimately be passed on to subsequent consumers). Furthermore, antitrust law treats buying cartels and horizontal mergers that create monopsony power similarly to the way it treats price-fixing and the creation of market power by sellers.¹⁴¹ Yet lower input prices can, to a degree depending on market structure and other factors, translate into lower prices to final consumers. If only consumer welfare mattered, increases in buyer power through horizontal mergers and otherwise might be praised, not condemned.

Pragmatic considerations may, however, provide justification for cabining efficiency defenses in various respects, including perhaps a focus limited to consumer welfare. For example, adopting a consumer welfare standard may induce firms to undertake deals that obtain potential synergies while causing less harm to competition, leading to even higher total welfare than would a total welfare standard. In any event, the previously-noted difficulties of ascertaining merger-specific efficiencies in individual mergers, *ex ante*, counsels caution, which is already reflected in the Merger Guidelines and also seems evident in court decisions. Indeed, this may explain why average efficiencies heavily influence the thresholds for anticompetitive effects, both generally and in specific industries, while at the same time efficiencies are often given short shrift in examining particular cases, absent an unusually strong demonstration. But once sufficiently persuasive proof of atypically great efficiencies is offered, it is not clear that a requirement of

¹³⁸ See Areeda, Hovenkamp, and Solow (1998, vol. 4A, pp. 37-38 n.1).

¹³⁹ A total welfare standard is also consistent with maximizing worldwide welfare. When there are multiple jurisdictions with authority over mergers with international spillovers, different jurisdictions might find it in their national self-interest to adopt different standards. For example, a country that is weighted toward consumers in sectors most influenced by antitrust policy might want to be stricter and use a consumer welfare standard, whereas a country with stronger producer/owner representation among its citizens might prefer total welfare or even a focus on productive efficiency. See Guzman (1998). Some have suggested that a mix of national bias and protectionism—rather than (or in addition to) differences in underlying approaches—may explain the (few) instances (the most notable being the proposed but ultimately abandoned merger of G.E. and Honeywell and the ultimately approved merger of Boeing and McDonnell-Douglas, subject to some conditions) in which the U.S. and EU antitrust authorities have taken different views of mergers or other practices.

¹⁴⁰ See, for example, Kaplow and Shavell (1994) and Kaplow (2004).

¹⁴¹ For example, the Merger Guidelines state: “The exercise of market power by buyers (‘monopsony power’) has adverse effects comparable to those associated with the exercise of market power by sellers. In order to assess potential monopsony concerns, the Agency will apply an analytical framework analogous to the framework of these Guidelines.”

pass-through to consumers makes the agencies' or courts' task easier rather than harder. On one hand, determining pass-through requires resolution of an additional, challenging issue. On the other hand, in some instances the best evidence of a merger's effects will be from observed pricing behavior that may reflect a combination of anticompetitive effects and efficiencies. For example, in *Staples*, evidence of higher prices in more concentrated markets presumably reflected the combined impact of concentration and whatever pass-through of efficiencies may have been occurring. Note that reductions in marginal cost are generally passed through to some degree to consumers, even by a monopolist, although the pass-through rate can be sensitive to the shape of the demand curve and the nature of oligopolistic interaction.¹⁴²

There are also longer-run concerns relating to dynamic efficiency. Some degree of competition has often been thought conducive to firms' running a tight ship, better serving customers, and being more innovative. The relationship between competition and innovation has proven difficult to establish as a general theoretical matter.¹⁴³ In any event, if the process of competition itself—which seems to have been favored by the drafters of antitrust legislation and earlier courts and commentators—is of some value, but this value is difficult to measure, it makes sense to tilt the balance against concentration. This might be done by making the threshold for challenge lower—requiring less demonstration of anticompetitive effects—or through other means, such as being less generous in considering efficiencies in justifying otherwise problematic mergers. Greater stinginess might be accomplished by raising proof burdens or by imposing additional requirements, such as by requiring savings to be passed along to consumers. The optimal manner of incorporating these sometimes subtle and typically unpredictable dynamic concerns into horizontal merger policy is not obvious.

E. Market Analysis Under the Horizontal Merger Guidelines

The Merger Guidelines are utilized to evaluate specific proposed mergers, requiring one to move from abstractions and general theories to decisions about a concrete case. To motivate this problem, consider the recent proposed merger of US Airways and Delta Air Lines. In November 2006, US Airways made an unsolicited offer to acquire Delta, which at the time was in bankruptcy. Predicting the effects of this merger on airline passengers is highly complex. First, one must identify the routes on which US Airways and Delta are significant direct rivals. Then one seeks to determine whether there will be a unilateral or coordinated price increase on those routes. Diversion ratios and gross margins are certainly relevant to this inquiry, but many other factors enter into the picture as well. Will competition from other carriers, including low-cost carriers, prevent the merged entity from profitably raising fares? Given the complex fare structures that are used in the airline industry, will the merger have different effects on fares for different classes of customers, such as leisure versus business travelers? How do frequent-flier programs affect the analysis? Will the merger generate substantial efficiencies based on running

¹⁴² Bulow and Pfleiderer (1983) derive an expression for the pass-through rate for a monopolist. Shapiro (1989) discusses pass-through rates in various oligopoly models.

¹⁴³ Competition generally enhances innovative incentives, as firms seek to gain ground on their rivals by introducing new and improved products. But a competitive market structure can cause problems for innovation if firms have difficulty appropriating the returns from their innovative efforts, for example, due to rapid imitation by rivals.

a larger network of flights at a single airline? In evaluating these efficiencies, how does one factor in the role of airline alliances, a less complete form of collaboration than a merger? If efficiencies lead to lower fares on some routes, but if the reduced competition leads to higher fares on other routes, how does one balance these diverse effects on different sets of consumers? Since one is looking ahead a year or more, is the industry changing in significant ways, such as through the growing role of regional jets and low-cost carriers, that make it appropriate to discount historical experience when making predictions? Lastly, how strong a competitor will Delta be if it is not acquired by US Airways, especially given the frequency of bankruptcies in this industry?

In this Section, we will consider only some of the core, recurring issues.¹⁴⁴ The first step in the analytical framework provided by the Merger Guidelines is defining the relevant market, typically a relevant product market along with a relevant geographic market. Then, as previously described in subsection D.2, the government determines whether there is a likely competitive risk based on the post-merger HHI and the increase in HHI due to the merger. If there is, it considers entry and other forms of supply response (as well as efficiencies, already discussed in subsection D.3, and the possibility that one of the firms is failing). In this section, we describe this approach and relate it to the preceding economic analysis and further work.

1. Market Definition: General Approach and Product Market Definition

In evaluating methods used to define the relevant market, one should ask whether the resulting measures of concentration are reasonably probative of anticompetitive effects in instances in which concentration is high and notably increased by the merger. This question is particularly pressing because, as we explained in subsection II.D.2.b, the common approach of defining markets and looking to market shares is just one of many, not necessarily the most reliable, and rather indirect and incomplete—indeed, sufficiently so that it does not play a central role in economists’ analysis of market power.

It is useful to begin by revisiting the precise relevance of market definition and market share. A single-product firm’s pricing decision is governed by equation (1),

$$\frac{P - MC}{P} = \frac{1}{|\epsilon_F|}$$

which is to say that the firm’s margin is inversely related to the magnitude of its own, firm-specific elasticity of demand. In the model of a single, dominant firm pricing with a competitive fringe, this firm-specific elasticity is given by equation (2),

$$|\epsilon_F| = \frac{|\epsilon_D| + (1 - S)\epsilon_R}{S}.$$

¹⁴⁴ Some issues that we do not cover, such as whether Delta is appropriately considered a “failing firm,” are treated extensively in the Merger Guidelines. Most others are noted in passing or would be admissible in light of various catch-all phrasings inviting consideration of any pertinent factors.

Thus, we recall that the firm's market share, S , is relevant for two reasons. First, a higher share means that there are fewer competitors, so for a given elasticity of supply response, the total response will be smaller. (Hence the $1 - S$ in the numerator.) Supply response will be considered further in subsection 3. Second, a higher share indicates that the firm captures a greater proportion of the industry profits due to a price increase.

Expression (2) also indicates, as we emphasized previously, that the firm-specific elasticity of demand depends on the market elasticity of demand, which raises complex problems for an approach that first defines a market and then looks to market share. If the market elasticity of demand were the same in every properly defined market, then (setting aside issues concerning rivals' supply response) shares would have clear implications for market power. However, this is certainly not the case (and, if it were, no one has ever suggested what that common market elasticity is). Generically, an approach that defines the relevant market (whatever that might mean) and then looks only at market share can be highly misleading. As we explained, a high share does not imply substantial market power if market demand is highly elastic, and a low share does not imply a lack of significant market power if market demand is sufficiently inelastic.

At best, markets can be defined so as to minimize these concerns, although as we explained, one often needs to know the right answer—that is, how much market power exists—in order to know which market definition is best. A major factor that contributes to this difficulty is the all-or-nothing nature of market definition: products (or regions) are either “in” or “out.” Moreover, we discussed how even good substitutes, in the sense of having a high cross-elasticity of demand with the firm's product, may impose little restraint, notably, if they are a small share of consumers' expenditures, whereas a group of mediocre substitutes might, taken together, impose substantial restraint.

Nevertheless, the Merger Guidelines in fact utilize an approach that relies heavily on defining markets, and the outcome of litigated merger cases often turns on how the relevant market is defined. Accordingly, it is important to consider how this task is presently accomplished and how it might be improved. For the case of product markets (the analysis of geographic markets is analogous), the Merger Guidelines (§1.11) specify the following procedure:

Absent price discrimination, the Agency will delineate the product market to be a product or group of products such that a hypothetical profit-maximizing firm that was the only present and future seller of those products (“monopolist”) likely would impose at least a “small but significant and nontransitory” increase in price [SSNIP]. . . . Specifically, the Agency will begin with each product (narrowly defined) produced or sold by each merging firm and ask what would happen if a hypothetical monopolist of that product imposed at least [an SSNIP], but the terms of sale of all other products remained constant. If, in response to the price increase, the reduction in sales of the product would be large enough that a hypothetical monopolist would not find it profitable to impose such an increase in price, then the Agency will add to the product group the product that is the next-best substitute for the merging firm's product. . . . The price increase question is then asked for a hypothetical monopolist controlling the expanded product group. . . . This process will continue until a group of products is identified such that a

hypothetical monopolist over that group of products would profitably impose at least [an SSNIP], including the price of a product of one of the merging firms. . . . In attempting to determine objectively the effect of [an SSNIP], the Agency, in most contexts, will use a price increase of five percent lasting for the foreseeable future.¹⁴⁵

This approach focuses on the ability of a hypothetical monopolist to raise prices, but the question at hand is a merger, and typically not a merger to monopoly. The behavior of the hypothetical monopolist seems most relevant when considering the possibility of coordinated effects. In the theory of collusion that we discussed in Section III, we implicitly assumed that we knew the group of firms in the market that might collude. If, regarding the merger under examination, that group of firms corresponds to the firms in the market just defined, then the Merger Guidelines approach will (supposing for the moment that it works) have indicated that successful collusion by the firms in the industry would indeed have a significant anticompetitive effect. If not—if, say, the market properly includes firms producing products very different from those of the merging firms, where heterogeneity or other factors are such that collusion is unlikely to be feasible—then the analysis would suggest that coordinated effects are probably not a concern. The second question with coordinated effects is, assuming that collusion would be profitable, how much does the present merger increase the likelihood of successful collusion? We examined this question in Section B, above, where we discussed various reasons that higher concentration and greater increases in concentration bolster collusion but also noted some reservations (namely, that if collusion is sufficiently likely short of monopoly, it is possible that further increases in concentration would not have much incremental anticompetitive effect).

The role of market definition is less clear, however, under a theory of unilateral effects, where the profitability of a price increase for the merged entity depends on the demand system and on such factors as the gross margins and the diversion ratios that measure the proximity of the merging firms' products, not on drawing lines between products that are in or out of the relevant market. However, in the particular case in which there is a logit demand structure, the elasticity of demand for the inside products as a group is relevant to calculating unilateral effects. If the merged entity would have only a small share of the relevant market, calibrated and defined using the SSNIP test, unilateral competitive effects in excess of the amount used to define the SSNIP are relatively unlikely, but such effects may well arise if the firms have a large combined share in the relevant market. Defining markets in this way thus allows the government to meet its presumption under a unilateral effects theory if the combined share of the merging firm exceeds

¹⁴⁵ When price discrimination is feasible and profitable, the Merger Guidelines (§1.12) allow for separate “price discrimination markets” “consisting of a particular use or uses by groups of buyers of the product” for which an SSNIP would be profitable. For example, in a railroad merger, the hypothetical monopoly rail carrier on a given route may be able to price discriminate between shippers who have the ability to ship by water versus those whose next best alternative to rail service is trucking. See Varian (1989) on price discrimination generally and Stole (2007) on its relationship to imperfect competition. In some circumstances, the hypothetical monopolist may be able to engage in price discrimination even if such discrimination is not observed prior to the merger because it is undermined by competition among the suppliers in the proposed relevant market. However, Hausman et al. (1996) point out that price discrimination can be unprofitable if the methods used by the hypothetical monopolist to price discriminate are imperfect.

some threshold.¹⁴⁶ Also, in industries where products are fairly homogeneous and capacities are important, such as in some chemical and energy markets, using the Merger Guidelines to define a relevant product and geographic market, and making inferences about unilateral effects based on concentration measures, fits reasonably well with the theory of Cournot equilibrium.

Now consider how, as a practical matter, one might implement the Merger Guidelines' approach to defining markets. For simplicity, consider a symmetric situation in which the pre-merger price of each product is P , the pre-merger output of each product is X , and the marginal cost of producing each product is a constant, MC , so the pre-merger gross margin is $m \equiv (P - MC) / P$. In this case, the SSNIP test reduces to a very simple formula. Suppose that the hypothetical monopolist raises the price of one of the products by a factor G , to $P(1 + G)$. (The same calculation could be done for a price increase applied to all of the products.) If the price increase would cause no decline in unit sales, it obviously would be profitable, whereas if it would cause unit sales to drop to zero, it would not be profitable. Since profits are continuous in sales, there exists an intermediate level of sales for which the price increase would break even. The largest percentage reduction in sales such that this price increase is just barely profitable is referred to as the "critical loss," L . By definition, $(P - C)X = (P(1 + G) - MC)X(1 - L)$. Solving for L gives

$$L = \frac{G}{G + m}$$
. This expression comports with intuition: a larger loss of sales is tolerable if the price increase is greater, but the acceptable loss of sales is smaller, the larger is the initial profit margin. For illustrative purposes, suppose that the magnitude of the SSNIP is 10%, so we are considering $G = 0.1$, and that the pre-merger margin is 30%; then $m = 0.3$ and we get $L = 0.25$.

These calculations are directly relevant to the SSNIP test in the Merger Guidelines. Consider a cluster of products that is being tested to see if the group of products are sufficiently inclusive to form a relevant market. Suppose that the pre-merger gross margin on each product is the same and, furthermore, that one is asking about a uniform price increase for all of these products and all customers. That hypothesized SSNIP will be profitable, and hence the products will form a relevant market under the SSNIP test, if and only if the actual loss of sales is less than the critical loss. Thus, in the foregoing example above, the hypothetical monopolist would find it profitable to impose a 10% price increase so long as the resulting actual loss would be less than 25% of the initial level of sales. Using simple arithmetic, we have thus translated the SSNIP test into a very well-defined economic question regarding the (arc) elasticity of demand facing the hypothetical monopolist. (Keep in mind, however, that this analysis is applicable to a uniform price increase in all products in the market, not to whether two merging firms would find it profitable to unilaterally increase the prices on one or both of their products.)

¹⁴⁶ The Merger Guidelines provide that "Where market concentration data fall outside the safe harbor regions of Section 1.5, the merging firms have a combined market share of at least thirty-five percent, and where data on product attributes and relative product appeal show that a significant share of purchasers of one merging firm's product regard the other as their second choice, then market share data may be relied upon to demonstrate that there is a significant share of sales in the market accounted for by consumers who would be adversely affected by the merger."

Katz and Shapiro (2003) and O'Brien and Wickelgren (2003) show how one can go beyond the simple arithmetic underlying critical loss to use pre-merger equilibrium relationships to sharpen the SSNIP test. Their work is motivated by the observation that the critical loss falls if the margin is higher. In the example just given, if the gross margin rises from 0.3 to 0.4, the critical loss falls from 25% to 20%. Based on this observation, merging parties whose products are sold at large gross margins have been known to argue for the inclusion of many products in the relevant market, based on the fact that the critical loss is small. However, these articles show that such logic is incomplete and can be highly misleading. The reason is that the existence of high pre-merger margins itself indicates that the pre-merger firm-specific elasticities of demand must be relatively small, and this of course has direct implications for the elasticity of demand facing the hypothetical monopolist in the post-merger market.

To see how the logic works, consider imposing the percentage price increase G on one product in the symmetric case just discussed. Since the firm-specific elasticity of demand for that product is the inverse of the margin, the percentage of sales lost for this product will be approximately G/m when G is small. Katz and Shapiro define the aggregate diversion ratio D for a given product as the fraction of the overall sales lost by that product that are captured by (diverted to) any of the other products in the candidate product market.¹⁴⁷ Therefore, the hypothetical monopolist who raises the price of this product effectively only loses a fraction $(1 - D)$ of the sale that are lost by that particular product. As a consequence, the actual loss of sales for the hypothetical monopolist is only $A = (1 - D)(G/m)$. The price increase is profitable if and only if the actual loss is less than the critical loss of $L = G/(G + m)$. With a few steps of algebra, it can be shown that $A < L$ if and only if $D > L$.¹⁴⁸

Where applicable, this formula tells us that a group of products will form a relevant market when the aggregate diversion ratio is larger than the critical loss. To illustrate, suppose that the pre-merger gross margins are 40% and that one is considering a 10% SSNIP, so the critical loss is 20%. Now imagine that the price of one of the candidate group of products is raised. If more than 20% of the lost sales are diverted to other products in this group, rather than outside the group, the group forms a relevant market. Katz and Shapiro (2003) stress that this test can lead to relatively narrow markets, especially in cases where the pre-merger margins are large. O'Brien and Wickelgren (2003) argue that critical loss analysis has often been done incorrectly, in a way that is inconsistent with the pre-merger equilibrium conditions, which they illustrate

¹⁴⁷ In the symmetric case with N products, the aggregate diversion ratio is equal to $N - 1$ times the diversion ratio between the product in question and any one of the other products. More generally, the aggregate diversion ratio is the sum of the diversion ratios for all of the other products being considered.

¹⁴⁸ Katz and Shapiro (2003) show how the calculations change if the pre-merger margins differ for the different products in the candidate market. In such cases, the Merger Guidelines ask whether the hypothetical monopolist will raise the price of *any* of the products sold by the merging firms. It may be profitable to raise the price of a product with an especially low margin because diversion of sales to other products will actually raise the profits of the hypothetical monopolist, offsetting sales that are not diverted to other products in the candidate group.

with two litigated merger cases, *FTC v. Tenet Healthcare Corp.*, a hospital merger, and *FTC v. Swedish Match*, a merger involving loose leaf chewing tobacco.¹⁴⁹

2. Geographic Market Definition

The Merger Guidelines approach to geographic market definition formally parallels its approach to product market definition. To illustrate how the analysis would proceed, consider BP's acquisition of Arco. These two companies were the leading producers of Alaskan North Slope crude oil, the vast majority of which was sold to refineries on the U.S. West Coast. These refineries required crude oil to make refined products, notably gasoline, so the relevant product was clearly crude oil. To determine the geographic market for the supply of crude oil to U.S. West Coast refineries under the Merger Guidelines, one starts with the merging firms' production locations, Alaska. One asks whether a hypothetical monopolist over Alaskan crude oil could profitably impose an SSNIP, taking as given the price of crude oil supplied from other locations to the West Coast refineries, which in this case included California and a variety of countries from which crude oil was shipped to the U.S. West Coast by tanker. If substitution to California crude oil would defeat a price increase imposed only on Alaskan North Slope crude oil, then California must be added to Alaska and the exercise repeated for foreign sources. In this case, since many West Coast refineries had demonstrated the ability to shift to imported crude oil in response to small price changes, the relevant market was arguably worldwide.¹⁵⁰

For the purposes of geographic market definition, it is common to look at the patterns of imports and exports across a given geographic boundary. However, care must be taken, for otherwise this method can be misleading.¹⁵¹ A prominent way of examining imports and exports is the method advanced by Elzinga and Hogarty (1973), which has been widely used in recent years in hospital merger cases.¹⁵² Basically, one starts with a narrow geographic market (typically a

¹⁴⁹ *FTC v. Tenet Healthcare Corp.*, 17 F. Supp. 2d 937 (E.D. Mo. 1998), rev'd, 186 F. 3d 1045 (8th Cir. 1999); *FTC v. Swedish Match*, 131 F. Supp. 2d 151 (D.D.C. 2000).

¹⁵⁰ See Bulow and Shapiro (2004) and Hayes et al. (2007) for further analysis of geographic market definition in this case.

¹⁵¹ One might wonder how it can be widely recognized that nontrivial product substitution is consistent with the existence of significant market power whereas geographic market substitution (imports and exports) is so often viewed as strong presumptive, if not conclusive, proof of the lack of market power. The common reasoning seems to imagine a case of homogenous products where the only distinction involves transportation costs, which are identical for all consumers at a given location. Think about supplies of fungible intermediate goods to firms. If a fraction of local firms is already importing its supply, all local producers must be indifferent between local purchases and imports at current prices, so an increase in local prices would cause substitution limited only by the extent to which foreign supply net of foreign demand slopes upward. But this reasoning is inapplicable when products are not homogeneous or consumers' locations or transportation costs vary (the latter of which is often true when the consumers rather than the goods must travel). Then, the relevant analogy is to markets with differentiated products, as will become apparent in the discussion to follow in the text. See, for example, Kaplow (1982).

¹⁵² See Frech et al. (2004) for an extensive discussion of how these tests have been used in hospital merger cases. Capps et al. (2002) emphasize problems with using patient flow data to define markets in hospital mergers and suggest an alternative approach based on estimating a logit model of hospital demand. They find that mergers that would be permitted by relying on patient-flow data may easily fail according to the SSNIP criterion under all three approaches that they examine; indeed, such data were largely uncorrelated with SSNIP except in extreme cases.

modest radius around the locale of two merging hospitals) and then progressively expands the candidate geographic market to include more distant hospitals until two conditions hold: (1) some large fraction of the merging hospitals' business comes from the candidate geographic market, and (2) some large fraction of the individuals resident in the candidate geographic market use hospital services within that region.

While such measures of imports and exports can be informative for the hypothetical-monopolist SSNIP test, and for estimating the competitive effects of mergers, they cannot substitute for evidence regarding the ability and willingness of consumers to substitute products supplied outside versus inside a candidate geographic market in response to price changes.¹⁵³ To see the problem, consider the case in which the two merging hospitals are nearby, with no other hospitals in the immediate area. Unless the fraction θ of the customers served by these hospitals that comes from the local area is very high, the proposed approach would draw the market more broadly.¹⁵⁴ How does this reasoning relate to whether an SSNIP could profitably be imposed by the merged entity, which is in fact a monopolistic supplier in the local area? Suppose that the elasticity of demand from local customers is ε_L and the elasticity of demand from customers who must travel further is ε_T . If the local customers can be identified and discriminated against by the merged entity, the fraction of customers coming from outside the local area, $1 - \theta$, is of little or no relevance to an SSNIP targeted at local customers, so there is no reason to believe that the test gives a meaningful answer. How does this test perform when price discrimination is not possible, and when more distant customers exhibit more elastic demand? The elasticity of demand facing the hypothetical monopolist is given by $\theta\varepsilon_L + (1 - \theta)\varepsilon_T$. With $\varepsilon_T > \varepsilon_L$, this is decreasing in θ , that is, the elasticity is greater, the more patients come from outside the local area. But this hardly answers the SSNIP question, which requires information about the pre-merger margins, the critical loss, and the actual loss, which depends on $\theta\varepsilon_L + (1 - \theta)\varepsilon_T$. In fact, the methods of Katz and Shapiro (2003) and O'Brien and Wickelgren (2003) discussed above tell us that the local area is a local market if and only if the diversion ratio from one hospital to the other is greater than the critical loss. These measures, in turn, bear no necessary relationship to measures of the fraction of demand coming from local residents or the fraction of services local residents obtain outside the proposed narrow market. That is, a local geographic area may well constitute a relevant geographic market, under the SSNIP test, even if a significant number of patients served by the local hospitals come from outside the area and even if a significant number of patients who live locally use more distant hospitals.

¹⁵³ Whinston (2006, pp. 92-93) comments on the "serious flaws" that can arise by defining relevant markets based on transshipment patterns.

¹⁵⁴ In practice, this fraction will depend on the type of service. For example, markets for emergency care or maternity care may well be more local than markets for elective surgery. The geographic market definition exercise must, in principle, be performed for each separate relevant product, at least if these products are priced separately, although sometimes the focus is on the average for all services, which is systematically misleading. (In our example, it may be that a high fraction who have unusual or extreme conditions travel a good distance, to major medical centers, whereas travel would not be worthwhile for most other medical needs.) Another practical difficulty is that consumers' locations are usually measured from the zip code that appears in admission records, but some individuals live part of the year elsewhere or they may work far from their residence, so for them a more distant facility may be less inconvenient (or may be the only one available under their employer's health plan).

Looking at import and export data can also readily lead to the opposite error: incorrectly concluding that a local region is a relevant market. Recall our example in subsection II.D.2.b of a trucking company that served 100% of a particular route, yet a slight price increase would lead nearby trucking companies to offer their services. To consider a more substantial example, return to the BP/Arco merger and ask whether Alaska and California constitute a relevant geographic market for crude oil. As long as transportation costs are not essentially zero and product differentiation is minimal, one would expect oil supplies to flow to the nearest refineries. Suppose that the supply of crude oil from Alaska and California is sufficient to supply 95% of the needs of refineries on the U.S. West Coast and that more than 95% of crude oil from Alaska and California is used at West Coast refineries. The Elzinga-Hogarty tests would indicate that Alaska and California qualify as a relevant geographic market. But this result would be incorrect if there is an elastic supply of crude oil from other locations at the current price, or at a price just above the current price but less than the price after the SSNIP, which would be true if the costs of importing oil from more distant locations were not significantly greater than that of transporting oil from Alaska. In fact, there is strong evidence that the supply of imported crude oil to the U.S. West Coast is highly elastic, with prices set by transportation arbitrage conditions.

3. Rivals' Supply Response

As we noted (reminded) in subsection 1, the ability to raise price depends not only on market share—and on the market elasticity of demand—but also on rivals' supply response, the subject of subsection II.D.2.c. The Merger Guidelines explicitly take this dimension into account. They distinguish between “uncommitted entry” and “committed entry.”

Uncommitted entrants are firms that would likely enter the market in response to an SSNIP in less than one year and without the expenditure of significant sunk costs. They are counted as market participants. That is, in computing market shares, their capacity is taken into account.¹⁵⁵ For example, a firm that currently makes few sales in the relevant market would be assigned a large share if that firm would greatly expand its sales in response to an SSNIP. By defining market share in this manner, the supply responses of all market participants to an SSNIP are incorporated into the analysis. This approach helps to adjust for the fact that the Merger Guidelines define the relevant product and geographic markets solely on the basis of demand-side considerations. If the government's *prima facie* case is established using these shares in markets thus defined, there is a presumption that current competitors' ability to expand is not sufficient to alleviate concerns associated with the merger.

However, it is also possible that the prospect of additional, committed entry, which might take as long as two years and require the expenditure of costs that then would be sunk, will either deter any post-merger price increase or insure that it is short-lived. In order for the anticompetitive

¹⁵⁵ There are additional subtleties involved in computing firms' market shares. In some cases, especially where products are differentiated, market shares are based on sales. In other cases, capacities are used. The general principle is that “Market shares will be calculated using the best indicator of firms' future competitive significance.” For example, to the extent that a firm's capacity is committed—perhaps under a long-term contract, or perhaps to other highly profitable uses, including those within a vertically integrated firm—and thus unavailable in response to an SSNIP, that capacity will not be included in measuring the firm's market share.

effects of a highly concentrating merger to be negated by the prospect of entry, the Merger Guidelines require that such entry be timely, likely, and sufficient to counteract the competitive effects of concern. Werden and Froeb (1998) explain how entry may well not be profitable if the pre-merger equilibrium reflects the sunk costs associated with entry. They also show that, if entry does occur, it can make mergers unprofitable unless they generate synergies. Baker (2003b) discusses the role of entry in recent merger cases.

F. Predicting the Effects of Mergers

In practice, myriad factors can come into play when predicting the competitive effects of mergers. Every industry has its own unique attributes, be they scale economies, the role of advertising and reputation, networks effects, consumer switching costs, product compatibility, technological change, regulations, intellectual property rights, or trade barriers. Virtually any topic in industrial organization economics can come into play in merger analysis. We confine our attention here to the two leading techniques used to predict the price effects of mergers, with the caveat that to use these methods reliably in any given case one must have a thorough awareness of industry-specific factors.

1. Direct Evidence From Natural Experiments

In some cases, a reduced-form approach is possible, under which empirical evidence is presented showing directly, through “natural experiments,” that the merger will lead to higher prices. The *Staples* case is an excellent example of the successful use of this approach by the government.¹⁵⁶ In this case, the FTC successfully challenged the proposed merger between Staples and Office Depot, two of the three leading office superstore chains (the other being OfficeMax). As explained by Baker (1999), the FTC presented econometric evidence showing that the prices charged by Staples at stores facing competition from nearby Office Depot stores were lower than the prices charged by Staples at stores without nearby Office Depot stores. The FTC offered further expert testimony showing that these price differences were not caused by other factors that happened to be correlated with the presence or absence of nearby Office Depot stores.

Economists might say that the FTC’s reduced-form approach, by going directly to the likely effects of the merger, reduced or even eliminated the need to define the relevant product market. In practice, however, this evidence was used in no small part to establish that “the sale of consumable office suppliers through office superstores” was a relevant product market, rather than the broader market for retail sales of office supplies, as alleged by the merging parties.¹⁵⁷

To take another example, in the Union Pacific/Southern Pacific railroad merger, the DOJ opposed the deal, arguing that freight rates would rise as the number of carriers on many routes declined from 3 to 2 or even 2 to 1. The DOJ supported its claims by presenting cross-sectional regression models showing how freight rates varied with the number of carriers on a route, along with other factors. Based on these reduced-form estimates, freight rate increases of about 20%

¹⁵⁶ Federal Trade Commission v. Staples, Inc., 970 F. Supp. 1066 (D.D.C. 1997).

¹⁵⁷ See Dalkir and Warren-Boulton (2004) for further economic analysis of the *Staples* case.

were estimated on routes going from 2 to 1 carrier, and freight rate increases of about 10% were estimated on routes going from 3 to 2 carriers.¹⁵⁸

More generally, reduced-form methods ask about the relationship between market structure (such as market concentration or the presence of certain companies) and prices or other measures of market performance, without specifying a structural model of the market. These methods require variation in the data on market concentration or competition, and to obtain reliable results one must be careful to correct for other factors that may influence price. Baker and Rubinfeld (1999) provide a broad discussion of the use of reduced-form estimates to predict the effects of horizontal mergers.

2. Merger Simulation

A small industry has arisen in recent years that uses simulation methods to estimate the effects of mergers in markets involving differentiated products. Merger simulation has most commonly been employed to study mergers involving consumer products, for which highly disaggregated retail scanner data on prices and sales are often available. This approach to merger simulation is described in detail and surveyed by Werden and Froeb (2007). Epstein and Rubinfeld (2001, 2004) also provide a very useful discussion of the merger simulation methodology.

There are two steps to simulating mergers. First, a demand system for the differentiated products must be specified and estimated. Werden et al. (2004) emphasize that the weight given to merger simulations should depend on how well the specified model fits the industry, based on historical evidence.¹⁵⁹ Estimating a demand system for differentiating products can be highly complex and require a great deal of detailed data. A number of methods have been developed to limit the number of parameters that must be estimated. One is to build a model in which demand for the various differentiated products depends on their underlying characteristics, an approach pioneered by Berry (1994) and Berry, Levinsohn, and Pakes (1995), with application to the automobile industry. Nevo (2000a, 2001) applies similar methods in the ready-to-eat cereal industry, and Nevo (2000b) provides a practitioners' guide. Another approach is explored in Hausman, Leonard, and Zona (1994), who employ a multi-stage budgeting procedure, under which products in a market are sorted into sub-groups based on their characteristics and demand is then estimated using this additional structure. Werden and Froeb (1994) use a logit model, which imposes a great deal of structure but requires the estimation of relatively few parameters. Epstein and Rubinfeld (2001) advocate use of the "Proportionality-Calibrated Almost Ideal Demand System" (PCAIDS) model for calibrated demand simulation, which they consider superior to other calibrated-demand models, including the logit model.

Second, the post-merger equilibrium is simulated using the structural model that was fitted to pre-merger data. This involves solving the post-merger equilibrium conditions using the

¹⁵⁸ See Kwoka and White (2004) for a balanced presentation of this major merger case. They cite (pp. 40–41) industry evidence from 2001, several years after the merger, that shippers on the 2 to 1 routes were paying a 20% to 30% premium, consistent with the DOJ's estimates.

¹⁵⁹ For example, one can check to see if the estimated firm-specific elasticities of demand are consistent with the Lerner formula and separately observed measures of marginal cost.

parameters estimated based on pre-merger conditions. Note that this approach assumes that the same model of oligopoly, a Nash equilibrium in prices, applies before and after the merger, so the method is only capable of estimating unilateral effects, not coordinated effects. Peters (2006) evaluates the performance of these methods using data from the U.S. airline industry.

V. Monopolization

Whereas Section III addressed collusion—when a group of competitors act in the manner of a single firm—and Section IV examined horizontal mergers—when competitors join to form a single firm—here we analyze how competition policy limits the behavior of a preexisting single firm. As mentioned in Section II.E, the offense of monopolization under U.S. antitrust law has two requirements: monopoly power and exclusionary practices.¹⁶⁰ Accordingly, we begin by elaborating on the element of monopoly power from an economic perspective, drawing on our broader discussion of market power in Section II and its application to horizontal mergers in section IV.E. Then we consider antitrust law on monopolization, with regard to both monopoly power and the exclusionary conduct requirement, viewed generally. Finally, we examine the economics and law as applied to certain important practices, predatory pricing and exclusive dealing, in this relatively controversial realm of competition policy.¹⁶¹ Myriad additional practices, some the subject of substantial literatures, are not considered here, although some of the principles adduced in our discussions of predatory pricing and exclusive dealing are pertinent.¹⁶²

¹⁶⁰ EU competition policy regulates abuse of a dominant position, which is analogous to the anti-monopolization provision in U.S. law and will be mentioned below, mostly in notes in Section B.

¹⁶¹ Crandall and Whinston (2003) argue that significant consumer benefits did not result from the remedies ordered in a number of the most visible government enforcement actions under Sherman Act Section 2 (the monopolization provision) during the twentieth century. Even if these results are accepted, their approach does not tell us about the deterrence benefits that result from inducing changes in the behavior of monopolists. In a companion piece, Baker (2003a) reviews evidence of anticompetitive outcomes before the enactment of the Sherman Act, during the 1890-1910 period when enforcement of Section 2 was often ineffectual, and from other countries without comparable laws.

¹⁶² The practice of tying, under which a firm requires customers who are purchasing product A also to purchase product B, has received a great deal of attention in the economics literature and in the law. Nalebuff (2003) provides a clear and accessible explanation of the complex economic issues that arise when evaluating the effects of tying as well as multi-product discounts, also known as bundling. Tirole (2005) provides a practitioner-oriented introduction to some of the issues that arise in the area of tying. Whinston (1990) examines the strategic use of tying to foreclose competitors in an imperfectly competitive market for the tied good. Refusals to deal, under which a vertically integrated firm refuses to sell its upstream input to its downstream rivals, are another practice that has been studied extensively. Rey and Tirole (2007) examine a range of strategies of what they call “vertical foreclosure,” including the denial to rivals of access to a bottleneck input, as well as “horizontal foreclosure,” including bundling and tying. Katz (1989) surveys the literature on vertical contractual practices.

A. Monopoly Power: Economic Approach

1. Rationale for Monopoly Power Requirement

The monopoly power requirement under U.S. monopolization doctrine is that, as a prerequisite to liability, there must exist a significant degree of market power—how much is required will be considered in subsection B.1. The rationale for market power thresholds has been addressed previously. In Section II.E, we identified the screening function, that is, the reduction of false positives. The value of a significant market power screen is particularly important with regard to monopolization for two interrelated reasons: single-firm behavior (which is obviously ubiquitous) is being regulated, and it often is ambiguous whether that behavior is anticompetitive (and most is not). These points deserve some elaboration.

At the heart of a market economy is the principle that firms have free reign to compete aggressively to win business and earn profits, possibly vanquishing their rivals in the process. If one firm does gain a dominant position, that is the firm's just reward for best serving the interests of consumers. Imposing liability on companies that compete most effectively, perhaps to the point of driving their rivals out of business, would contravene the fundamental workings of a market economy. Furthermore, the argument goes, government intervention is unnecessary because even the most successful companies must continually face the gales of creative destruction, as new and innovative rivals challenge their positions. (And, in exceptional cases of natural monopoly, some form of industry-specific regulation is the answer, not broad-based limits on competitive practices.)

Even most who accept strong forms of this laissez-faire view would embrace rules against collusion and horizontal mergers to monopoly. A much greater danger, however, is raised when various practices of individual firms are challenged by the government—and, in some jurisdictions like the United States, private plaintiffs (often unsuccessful rivals). Here, even those highly skeptical of extreme laissez-faire recognize the potentially high costs of litigation, of erroneous condemnation of benign or affirmatively beneficial practices, and perhaps most importantly of chilling routine competitive behavior. As noted, the risks are especially great because it is often difficult to distinguish exclusionary from pro-competitive conduct, a subject to which we will return in Sections B-D.

Fortunately, for most firms in most industries, the danger of socially costly anticompetitive behavior is negligible because the firms lack significant market power or any serious prospect of acquiring it even using the challenged practices. Accordingly, for such firms, there is likely to be little benefit from examining in detail the effects of their conduct, whereas substantial costs of administration, mistaken prohibition, and inhibition of competitive vigor can be avoided by in essence granting them immunity. To give a concrete example, imposing a monopoly power screen in the area of predatory pricing avoids potentially enormous costs that could arise if every firm contemplating an aggressive low-price strategy had to fear a possible predatory pricing challenge from its rivals.

The monopoly power requirement is similar to the rule examined in subsection IV.D.2 that horizontal mergers must exceed some threshold level of anticompetitive effects as a prerequisite to liability. With mergers, some level of synergies are presumed to exist in typical cases, so

scrutiny is only triggered when anticompetitive effects are nontrivial. With monopolization, the threshold is ordinarily understood to be much higher, not so much because single-firm practices are ordinarily far more valuable than mergers, but rather because of the more substantial problem of false positives and related chilling effects.

Another point bearing on the value of a monopoly power requirement concerns the magnitude of incremental harm if prices do rise. If there is no technical market power as defined in Section II (that is, if price is at marginal cost), the marginal loss in total surplus as price begins to rise is zero. The greater the extent of initial market power—that is, the higher the initial margin—the greater is the marginal distortion from further price increases. Note, however, that this observation is pertinent under a total welfare standard rather than a consumer welfare standard, on which see subsection IV.D.3. After all, the marginal reduction in consumer surplus is highest when price rises from the point at which it equals marginal cost; the higher is the initial price, the smaller is the incremental reduction in consumer surplus. (The incremental reduction per unit of price increase is just the quantity demanded, which falls with price.)

Application of a monopoly power test raises a number of issues. First, how high should the market power requirement be? Unfortunately, this question is difficult to answer because it is so hard to measure most of the costs and benefits of a higher threshold, notably, the level of chilling effects and the relative proportions of beneficial and undesirable practices among those that are deterred. As we suggested in Section II.E, it seems optimal for the threshold to depend on the practice: for those obviously undesirable, little or no threshold seems necessary; for those more questionable, some intermediate standard; and for those sufficiently likely to be beneficial, an extremely high one—at some point being tantamount to deeming the practice legal *per se*. Again, to compare with horizontal mergers, many economies that may be realized through mergers can, although perhaps more slowly and at somewhat greater cost, be obtained through internal growth. But if innovation or aggressive, competitive pricing by individual firms is deterred, alternative outlets seem unavailable.

Another factor bearing on the height of the monopoly power threshold concerns the cost and potential for error in the market power inquiry itself. As discussed in Section II.D (and elaborated in Sections IV.E and IV.F with regard to horizontal mergers), there are numerous means of assessing a firm's market power that vary greatly across markets in their feasibility and reliability.¹⁶³ If a practice seems fairly clearly evil, it neither saves enforcement costs nor significantly reduces false positives to impose much (or any) market power requirement.

We also discussed how conduct itself may be highly probative of market power in cases in which the conduct would not be rational in its absence. Requiring proof of power without taking into account such conduct makes little sense, and if the conduct is, logically, used to infer sufficient power, then the monopoly power requirement is not serving as an independent threshold test. From an economic viewpoint, this is an appropriate result. The danger arises when the practice is more uncertain regarding whether it is desirable, is in fact being employed, or its use

¹⁶³ Indeed, much of what economics has to offer antitrust law concerns the assessment of market power. Because of our extensive treatment of the subject in these previous sections, we do not take up the matter further here.

necessarily implies the existence of significant market power. (These points are usefully reconsidered in the case of predatory pricing, taken up in Section C.)

2. Application to Challenged Practices

Subsection 1 addresses the purpose of a market power threshold in monopolization cases. We have not yet, however, revisited the important question that we raised in Section II.E about whether the requisite monopoly power is that which exists but for the challenged practices or in light of them, and what if any is the relevance of the difference between these two levels of market power. For horizontal mergers, recall from subsection IV.D.2 that the Merger Guidelines (in the United States, and similarly in the European Union) impose a two-part test that requires both that the post-merger HHI be sufficiently high and that the increase in HHI due to the merger be at least of a certain magnitude (which itself depends on the former measure). The analogue for monopolization would be to require that there exist monopoly power with the challenged practices and also that the practices contribute appreciably to that power. We defer discussion of the current state of the law to subsection B.1; here, we consider what economic analysis has to say about these issues.

At first glance, it might appear that only the increment should matter, for what is challenged is a set of practices, not the means by which the firm had previously gained its position. Thus, if a firm with a valid and powerful patent engages in a practice that slightly increases its monopoly power, ordinarily all that would be enjoined would be the illegitimate practice; likewise, fines or damages would be based on the addition to power, not the whole of profits legitimately attributable to the patent itself. Accordingly, it seems that the key question is not the disembodied query “How much economic power does the defendant have?” but rather “Will the challenged practices harm competition?” or “Will the challenged practices enhance the defendant’s market power?” That well-defined economic question often can be usefully recast as: “Will the challenged practices significantly remove or relax constraints on the defendant’s pricing?”

Our analysis in subsection 1, however, suggests that this should not be the sole inquiry regarding market power. The screening function and the related problems of false positives and chilling effects indicate that we probably should not freely allow challenges in industries where there is little market power or against particular firms with little power (although if the increment were sufficiently large, say, all the way to a monopoly, that would be another matter). Also, if prices are near marginal cost, price increases cause little deadweight loss.¹⁶⁴

Taken together, for many practices (setting aside those that are unambiguously undesirable), it may make sense both to insist that the firm possess some significant level of market power and that the challenged practices contribute importantly to it. Regarding the former, it often would not much matter whether the overall level of market power was measured with or without the challenged practices, unless they had a very large impact. In such cases, it probably makes sense to consider power with the practices, although if power is quite low without them and the

¹⁶⁴ Recall from Section II, however, that when there are significant fixed costs, equilibria will be characterized high margins, so the marginal welfare cost of a price increase would be nontrivial.

practices themselves are ambiguous, one may be skeptical of their effects and much of the screening function would be lost if extravagant claims were permitted against firms that had little market power. This assessment, however, imagines a prospective challenge against practices that have not yet had their (alleged) impact. If, instead, the practices have had time to take effect and the result is substantial market power, it hardly seems sensible to excuse the defendant that asserts its power would be small without the practices, for that would be an admission of large anticompetitive effects.

Consider further the notion that the challenged practices themselves should be shown to contribute significantly to the firm's market power. Just as in the case with horizontal mergers, if there were no possible benefits that might accompany the practices, it would seem that any increment (perhaps beyond a *de minimis* level) should be condemned. There are two important reservations to this conclusion in the monopolization context. First, as with horizontal mergers, some practices that may have anticompetitive effects may also promote efficiency. Consider, for example, the possible tradeoffs that may be involved with exclusive dealing, as we discuss in Section D. Then, one would need to balance the two, and how that balance should be conducted would importantly depend on whether the standard is limited to consumer surplus or is defined in terms of total economic welfare, as we discussed in subsection IV.D.3. Second, there is again the problem of uncertainty and false positives. If we are uncertain about whether challenged practices are undesirable at all, then depending on our Bayesian prior about the likelihood of different effects and the evidence before us, it may well be that the appropriate loss function is minimized by requiring that apparent anticompetitive effects be above some magnitude before condemning the practices.

Of course, applying this additional test regarding the increment to market power is not without cost because such an inquiry requires additional information about the conduct at issue. When practices are ambiguous—which is when screens tend to be most helpful—it will often be uncertain whether there is any anticompetitive effect, so one might wonder how its impact might be quantified. An answer is that sometimes it can be quantified conditionally; that is, one can assume that the practices have some specified type of effect and then attempt to quantify what that effect would be.

To apply these suggestions more concretely, it is useful to return to the framework introduced in Section II on market power and consider an example. In subsection II.D.2.b, when discussing substitutes, we derived equation (5), showing that the elasticity of demand for a given product (product 1 of the N products) is equal to one plus the sum of the cross-elasticities of all the other products with that product, each weighted by that product's share of expenditures:

$|\varepsilon_{11}| = 1 + \frac{1}{s_1} \sum_{i=2}^N s_i \varepsilon_{i1}$. According to equation (1), the profit-maximizing markup for a single firm

producing product 1 is given by $m_1 = 1/|\varepsilon_{11}|$. The expression for the elasticity of demand captures the familiar idea that the firm's market power is increased if the ability of consumers to shift to certain substitutes is reduced. Practices that reduce a number of these cross-elasticities ε_{i1} reduce the magnitude of the elasticity of demand for the product in question, $|\varepsilon_{11}|$, giving the firm greater market power and thereby raising the firm's profit-maximizing price.

Suppose that the practices at issue are alleged to reduce the attractiveness of some of the substitutes to the firm's product. For example, predatory pricing might eliminate one or more of the substitute products from the market. In principle, a full inquiry would allow us to determine that the practices reduce the cross-elasticity of demand with substitute i at preexisting prices from ε_{i1} to $\hat{\varepsilon}_{i1} \leq \varepsilon_{i1}$. The full inquiry thus would tell us that the practices reduce the magnitude of the elasticity of demand facing the firm at preexisting prices from $|\varepsilon_{11}|$ to $|\hat{\varepsilon}_{11}| = 1 + \frac{1}{s_1} \sum_{i=2}^N s_i \hat{\varepsilon}_{i1}$.

Since $|\hat{\varepsilon}_{11}| < |\varepsilon_{11}|$, the firm's profit-maximizing price will rise as a consequence of the practices.¹⁶⁵

Now return to our questions about levels of market power versus changes in power due to the challenged practices. Regarding the monopoly power threshold, if the magnitude of the elasticity of demand facing the firm without the challenged practices, $|\varepsilon_{11}|$, is sufficiently high, no liability can arise if the test looks to preexisting power. As noted, the rationale would be that a firm without significant market power initially cannot profitably create such power by engaging in the practices at issue, leading us to be skeptical about the alleged anticompetitiveness of the practice. Under this approach, if $|\varepsilon_{11}|$ is sufficiently high, one need not look at $|\hat{\varepsilon}_{11}|$. Alternatively, if the question is power with the practice—which is more natural to employ if the practice has existed for awhile—the question would be whether $|\hat{\varepsilon}_{11}|$ is sufficiently high.

We also considered the relevance of the extent to which the practices at issue may plausibly enhance the firm's market power, which focuses on the difference between $|\hat{\varepsilon}_{11}|$ and $|\varepsilon_{11}|$, not on the level of either in isolation. To see how this difference might be measured, suppose that the challenged practices only affect a certain subset of substitutes, J . At worst, the practice would eliminate those substitutes as alternatives available to consumers, which is equivalent to setting $\varepsilon_{1j} = 0$ for all $j \in J$. Without these substitutes, the magnitude of the firm's elasticity of demand

at preexisting prices would fall to $|\hat{\varepsilon}_{11}| = 1 + \frac{1}{s_1} \sum_{i \notin J} s_i \varepsilon_{i1}$; of course $|\hat{\varepsilon}_{11}| < |\varepsilon_{11}|$. If the gap $|\varepsilon_{11}| - |\hat{\varepsilon}_{11}|$

is small, so the challenged practices only modestly reduce the magnitude of the firm's elasticity of demand at preexisting prices, then these practices, even if effective, cannot lead to a significant increase in market power.¹⁶⁶

¹⁶⁵ Consumers would be harmed not only by the firm's price increase but also by any reduction in their utility from using the substitute products that results from the challenged practice. When we study exclusive dealing in Section D, we explain how a firm's conduct can make substitute products less attractive and whether there are offsetting benefits to consumers. Efficiencies could be included in this analysis by allowing the practices at issue to reduce the firm's marginal cost as well as its elasticity of demand, so the firm's price might fall even as its markup rises.

¹⁶⁶ This approach is analogous to the hypothetical-monopolist test that is used for horizontal mergers (see subsection IV.E.1) where we imagine that the hypothetical monopolist controls all of the products affected by the challenged practices.

B. Legal Approach to Monopolization

As we noted in Section II.E on market power, under U.S. Sherman Act Section 2 “[t]he offense of monopoly . . . has two elements: (1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident.”¹⁶⁷ Here we consider each in turn.

1. Monopoly Power

The central legal question is how much market power is denoted by “monopoly power.” The just-quoted authoritative statement from the Supreme Court does not answer this question. Unfortunately, even those cases that offer quantitative statements are far less illuminating than meets the eye.¹⁶⁸

Most famous is the pronouncement in *Alcoa* that a ninety percent share (in the market for aluminum) “is enough to constitute a monopoly; it is doubtful whether sixty or sixty-four percent would be enough; and certainly thirty-three per cent is not.”¹⁶⁹ The difficulty in interpreting this statement is that two distinct issues are conflated: how much market power was thought to exist in that case? (a fact question distinctive to that industry under the then-existing conditions), and how much market power is deemed sufficient to constitute monopoly power? (a legal/policy question, the answer to which may be entirely independent of the particular case or, if not, its dependence requires specification that was not offered). For example, might the court have thought that monopoly power consists of the ability to sustain a margin of at least 20%, that a 90% share conferred the power to price 35% above cost, a 33% share only 10% above cost, and a 60-64% share somewhere near 20%? Or might it have thought that monopoly power required only the ability to sustain a 10% margin, but that the power implied by each of the stated shares was only half as high?

This ambiguity is fundamental because in future cases, not in the aluminum industry (under the conditions prevailing at the time of *Alcoa*), a given share, whether 33%, 90%, or some other figure, may convey much more or significantly less power than did a similar share in *Alcoa*. But we know neither how much power over price *Alcoa* required nor how much power was thought to exist for any given share in that industry. Hence, even if both parties’ experts in a subsequent

¹⁶⁷ *United States v. Grinnell Corp.*, 384 U.S. 563, 570-71 (1966). We focus here on the antitrust violation of monopolization; attempted monopolization, which is noted briefly below, is similar but places less weight on the defendant’s current (versus prospective) monopoly position. We also note that in the European Union Article 82 prohibits “Any abuse by one or more undertakings of a dominant position,” which requires an analogue to both monopoly power (dominant position) and exclusionary conduct (abuse). As we will note at various points below, however, the interpretations have differed, although the divergence seems to be shrinking over time.

¹⁶⁸ For prior discussion and questions, see Areeda, Kaplow, and Edlin (2004, §3.B).

¹⁶⁹ *United States v. Aluminum Co. of America*, 148 F.2d 416, 424 (2d Cir. 1945). Although this was a Court of Appeals case, it was decided by a prominent panel of judges, the opinion was written by the famous Judge Learned Hand, the court was acting in lieu of the Supreme Court, and the opinion was blessed in subsequent Supreme Court cases.

case agree that, say, the sustainable margin was 16%, there is no way to tell from *Alcoa* which side wins on the element of monopoly power.

The same opacity characterizes all statements that a given market share is or is not adequate under any market power test—that is, unless one accepts that a given share in a properly defined market conveys the same market power, regardless of the market.¹⁷⁰ But this supposition is emphatically false. Instead, as developed in Section II.B and reemphasized in subsection IV.E.1 on market definition in horizontal merger cases, in the basic case a firm’s price-cost margin is the inverse of the firm-specific elasticity of demand, ε_F , which in turn is given by equation (2),

$$|\varepsilon_F| = \frac{|\varepsilon_D| + (1 - S)\varepsilon_R}{S}.$$

Thus, in addition to the market share S , both the market elasticity of demand, ε_D , and the elasticity of supply response, ε_R , are important, and we have seen that it is quite possible for a high share to be associated with low market power and a modest share to be associated with substantial market power.

Furthermore, as discussed at some length in subsection II.D.2.b, there is no means of defining markets in such a way as to circumvent this problem. Assuming some hypothetical, benchmark market in which a stipulated level of market power is associated with each market share, it would only be by chance that there would be a readily available market definition in any given case that would yield a share that indicates the correct amount of market power. This inability relates to the all-or-nothing nature of market definition—products or regions are either “in” or “out”—and the fact that the market elasticity of demand is, as noted in subsection A.2, one plus the (revenue-share) weighted sum of the cross-elasticities of all products, rather than being determined solely (or even primarily) by the cross-elasticities of one or two products. Note also that, even if one could match every actual market to such a hypothetical benchmark market, we have no way of knowing how the aluminum industry in *Alcoa* or other industries and markets in other antitrust cases relate to that imaginary market.¹⁷¹

Before proceeding, it should be noted that the foregoing problem does not bear on the utility of the standard practice of determining market power by defining a relevant market and then measuring the firm’s share in that market. As we discussed, consistent with the formula for a dominant firm’s elasticity of demand, these steps need to be supplemented. At best, courts tend

¹⁷⁰ In the European Union, there is an even greater tendency to rely on market share in cases alleging abuse of a dominant position, although if its merger law is any indication of a general trend, there seems to be a tendency to move toward a more economic approach to the assessment of market power, as we note in subsection IV.D.2. In any event, EU cases suggest that a 50% market share may well be enough, and for some practices even lower shares might be accepted.

¹⁷¹ As should be clear, the problem we identify is not unique to *Alcoa*. Consider any case that states, say, that a 50% market share is required. If that statement is to be associated with the market power that exists in the case at hand, then our discussion of *Alcoa* applies directly. If it is to relate to the level of power that exists in a “typical” market, we need to know what that market is and how much power is implied.

to do this indirectly. Thus, they attempt to define the relevant market such that the market elasticity of demand is not so great that high market shares in that market are consistent with negligible market power. Additionally, they check for the presence of some entry barriers, an aspect of possible supply substitution. Even if this is done well, however, and even if there is no better way to determine market power in a given case, one still needs to know how much market power is required, which is the question at hand.

Another problem with the failure to state the underlying market power threshold explicitly is that there is no way to relate the many other means of measuring market power that we examined in Section II.D to pronouncements about market share. In this regard, one can contrast the approach under the Horizontal Merger Guidelines under which “the Agency, in most contexts, will use a price increase of five percent lasting for the foreseeable future” to define the SSNIP. See subsection IV.E.1. Under such an explicit approach, it is possible to use the sorts of empirical techniques described in section IV.F to ascertain market power, such as in the *Staples* case where the government presented and the court was persuaded by evidence of price differences across regions with different numbers of competitors. As already noted, even if both sides were to agree on the level of market power in a monopolization case, we cannot tell from existing statements whether monopoly power would be deemed to exist. And since no quantitative threshold has been stated, there is no way of determining whether it reflects an appropriate balance of screening benefits, litigation costs, and so forth.

Consider next the relationship between the monopoly power requirement and the challenged practices. One issue is whether exclusionary conduct may be used as a basis for inferring monopoly power. Courts seem to contemplate that nearly any relevant evidence will be admitted but nevertheless are reluctant to find a violation unless the monopoly power requirement is established through proof of a relevant market in which substantial power exists. On one hand, it is rational to insist on proof of power—given the purpose of the monopoly power screen, to avoid false positives and related chilling effects—in cases when there is ambiguity about the practices under scrutiny. As we mentioned in a note in subsection II.D.3 on inferring power from conduct, the Supreme Court’s insistence in *Spectrum Sports* that a plaintiff in an attempted monopolization case must independently prove the requisite power may reflect the fact that the challenged act involved terminating the plaintiff distributor in favor of another, a practice that hardly evidences such power (although it seems to have convinced a jury).¹⁷² Requiring independent demonstration of a predator’s ability to recoup losses, to be discussed in subsection C.3, also seems to reflect skepticism about whether truly predatory pricing can confidently be identified. As already noted, however, if a practice is unambiguously exclusionary, there is reason to infer some market power, and there is little reason to impose a strong filter.¹⁷³

¹⁷² *Spectrum Sports v. McQuillan*, 506 U.S. 447 (1993).

¹⁷³ It is independently problematic that courts often seem to insist on defining a relevant market, which we have seen is only an aspect of one means of inferring market power—and not always the best means—although we also have noted that more recently courts have tended to accept more direct evidence when it is offered and found to be persuasive.

We also addressed whether it makes economic sense to focus on extant market power (whether with or in the absence of the challenged practices) rather than on how much practices contribute to that power. The monopoly power requirement itself seems focused on extant power. However, in showing that practices are exclusionary—on which more in subsection 2—it is required that they have contributed to that power. This demand is viewed as part of the determination of the second monopolization element rather than as part of the first. To illustrate the relevance of increments to market power, in the U.S. government’s settlement with Microsoft in the first wave of enforcement activity in the mid-1990s, the DOJ agreed to modest conduct remedies (which some commentators and intervenors viewed as too lax) precisely because DOJ insisted that only a moderate portion of Microsoft’s market power in operating systems was attributable to the challenged conduct.¹⁷⁴ The issue also arises in claims of attempted monopolization. To demonstrate the required dangerous probability of success, it is necessary to show that the practices will contribute appreciably to market power since the defendant in such cases is not yet imagined to have monopoly power.

In subsection A.2, we also briefly discussed whether monopoly power is to be gauged with or without the challenged practices. In many cases, the practices have been in place sufficiently long that the status quo plausibly reflects their effects (if any). This setting has sometimes led to confusion, most notoriously in the *du Pont (Cellophane)* case.¹⁷⁵ In defining a broad market—not just cellophane but also other flexible wrapping materials—the Supreme Court was heavily moved by the fact that many customers already used various alternatives to cellophane. As we explained in Section II.E, however, such a view implicitly asks whether the firm could profitably raise prices significantly above present levels. The answer to that question will almost certainly be negative even if power is great, for if higher prices were profitable, they would already be observed. Furthermore, if the firm-specific elasticity of demand is less than one, it is necessarily profitable to raise price, and to keep doing so until one hits a region of the demand curve that is more elastic. (The margin, given in expression (1), equals the inverse of the magnitude of the firm’s demand elasticity; since the margin is $m \equiv (P - MC) / P$, a finite P implies that the magnitude of the elasticity must exceed 1.) Since supply substitution was limited in *Cellophane*, it must have been that price was in an elastic region of the demand curve for cellophane; hence, profit-maximization implies that price must have been high enough that significant substitution occurred. This problem, the so-called “*Cellophane* fallacy,” reflects that, although courts have long viewed market power in terms of the ability to elevate price, they have only gradually incorporated economic analysis that bears on how the degree of market power is determined.¹⁷⁶ We also note that, to a substantial extent, the fault does not lie with the courts but rather with

¹⁷⁴ *United States v. Microsoft Corp.*, 56 F.3d 1448 (D.C. Cir. 1995).

¹⁷⁵ *United States v. E.I. du Pont de Nemours & Co.*, 351 U.S. 377 (1956).

¹⁷⁶ Consider also our discussion in subsection IV.E.2 of geographic markets concerning patient flows in hospital merger cases and imports more broadly. For example, with homogenous goods, common production costs, and positive transportation costs, the presence of imports implies that local producers’ mark-up equals transportation costs, which may be significant; the absence of any imports implies that the mark-up is less than transportation costs, suggesting that the elasticity of local demand or rivals’ supply exercises more of a pricing restraint. Hence, it would be a mistake to infer that the existence of imports—a readily observable form of substitution—implies less market power than would be implied by their absence.

litigants who have not made full, cogent use of economic teachings. In cases such as *Staples* in which courts have been offered direct and persuasive evidence and analysis, they often seem ready to accept it, even if their decision is still articulated using more traditional rubrics (in that instance, by defining a narrow market upon being presented evidence of sufficient market power).¹⁷⁷

2. Exclusionary Practices

Most economic analysis of exclusionary practices focuses on particular types of conduct, such as predatory pricing and exclusive dealing, which we take up in Sections C and D, respectively. Before considering specific applications, however, it is useful to consider the law's general formulation of the second element of a monopolization claim, with an eye to how it might be given an economic interpretation. Unfortunately, this aspect of monopolization law is also rather obscure.

The sort of authoritative statement with which this section began illustrates the problem. The quoted Supreme Court language refers to “the willful acquisition or maintenance of [monopoly] power as distinguished from growth or development as a consequence of a superior product, business acumen, or historical accident.”¹⁷⁸ The latter portion of this clause does clearly indicate that the term “exclusionary practices,” often used to capture this requirement, cannot be taken on its face, for superior products and business acumen tend to exclude inferior competitors from the market.¹⁷⁹ Thus, the test refers only to some subset of practices that exclude. But what subset? The quotation speaks of power acquired or maintained willfully, but that limitation is most unhelpful, for it suggests that only accidental or mistaken behavior is exonerated.

Another common formulation distinguishes between competition that is on the “merits” and that which is not. The reference to merits, however, is patently question-begging.¹⁸⁰ The use of the term competition is suggestive of the rule of reason's test of whether arrangements promote or suppress competition. As we discussed in subsection III.E.2, that test seems often (but not always) to suggest a process orientation toward the meaning of competition rather than a focus on the results of competition in terms of economic welfare. Given that the rule of reason is employed under Sherman Act Section 1 and that monopolization is prohibited by Section 2 of

¹⁷⁷ Another issue is the extent to which monopolization policy is guided by factors other than considerations of economic welfare (in total or that of consumers alone), a topic considered further in our discussion of the law on exclusionary practices that follows. This consideration is probably of greater contemporary importance under EU law on the abuse of a dominant position. See Hawk (1988).

¹⁷⁸ As elsewhere, our focus will be in U.S. antitrust law; the EU prohibition on the abuse of a dominant position suffers similar ambiguity. Additional specifications under EU law suggest a broader scope that, if taken literally, might include the ordinary monopoly behavior of elevating price and reducing output. Although the EU's prohibition does seem to be interpreted more broadly than Sherman Act Section 2, it is not given such breadth.

¹⁷⁹ The term is also under-inclusive. A horizontal merger by a firm with monopoly power may be considered to be an act of monopolization, but it would not ordinarily be characterized as exclusionary (although one could state that the formerly independent firm has thereby been excluded from the market).

¹⁸⁰ Similarly problematic are statements regarding the EU “abuse” requirement that it refers to other than “normal” competition.

the same Act, as well as the growing convergence of interpretation in the United States under all of its antitrust statutes, this possible connection between these two locutions is worth keeping in mind.

Commentators have recognized the ambiguity of the exclusionary practices requirement (both in the U.S. law of monopolization and the most closely corresponding EU law on abuse of a dominant position), and they have proposed a variety of ways to give meaning to the second element of monopolization charges.¹⁸¹ These alternatives are often of a more explicitly economic nature. Suggestions include a focus on consumer welfare, an inquiry into whether a firm has sacrificed short-term profits, an examination of whether a practice makes no economic sense but for its effect of excluding a rival, and an assessment of whether more efficient rivals are or can be excluded.

To obtain an overview of some of the choices involved in picking a general test, it is useful to consider briefly the core question of what price a monopolist is permitted to charge (anticipating somewhat our discussion of predatory pricing in Section C). To begin, it is well established that a monopolist is generally permitted to charge the classically defined monopoly price. Since this price tends to minimize both consumer and total surplus (relative to lower prices, that is), this broad permission seems inconsistent with (indeed, in contradiction to) an economic welfare standard.

There are two primary justifications for allowing monopoly pricing, and they provide insights into how one should think about the exclusionary practices requirement. First, monopoly profits often reward socially valuable *ex ante* investments, such as in innovation, cost-cutting, or generally running a tight ship. Thus, a dynamic view of welfare is adopted. Note, however, that this view is embraced broadly; defendants are not required to prove in particular cases of high pricing that their profits are efficient *ex post* rewards for prior good behavior. Second, price regulation is not thought to be in the institutional competence of courts or generalist antitrust regulators. To be sure, in certain cases (notably, natural monopoly), comprehensive price (and other) regulation is employed, but under the supervision of specialized agencies. These points are mutually reinforcing in that it is thought to be inefficient for inexpert institutions to intrusively interfere with single-firm behavior at the risk of dynamic efficiency. But the points are also in tension, for if the pertinent calculus is a complex, dynamic one, and the agencies and courts charged with the antitrust task have limited skills, how are they to identify which practices should be prohibited?

Having set aside the notion that a pure, general welfare test (whether total or consumer welfare) is to govern, some of the other proposed tests can be placed in context. Inquiring solely into whether there is a short-run profit sacrifice is obviously problematic, because any investment, whether in innovation, new plant, or even routine employee training, involves sacrificing short-run profits for long-term gain. Asking whether a practice is profitable but for its exclusionary effects is similarly deficient, for better products and superior service are profitable in significant part due to their tendency to capture business from rivals. (A firm researching a new cure would

¹⁸¹ See, for example, Areeda and Hovenkamp (2002, vols. 3, 3A), Elhauge (2003a), Melamed (2006), Popofsky (2006), Posner (2001, pp. 194-95), Salop (2006), Vickers (2005), and Werden (2006).

anticipate no revenue if no consumers would switch from existing treatments.) In addition, not all anticompetitive practices involve a short-term sacrifice.¹⁸² Banning only practices that keep out more efficient rivals might be consistent with maximizing productive efficiency but gives no weight to consumer surplus and accordingly would often permit practices that reduce total surplus as well.¹⁸³ The contrast is nicely posed by a case we will consider in Section C: if a firm's pricing strategy drives out (or deters) slightly less efficient entrants, with the result that prices are substantially higher and consumer and total welfare significantly lower, should a violation be deemed to occur? Better results may be possible by refining or mixing these tests, such as by insisting that a practice be profitable but for the possible additional price increment one can charge if a rival has been excluded from the market rather than hypothetically remaining and continuing to offer its product.

No simple solution is readily apparent. Any general test—or, in its absence, particular tests for particular practices—should probably be grounded in concerns for long-run economic welfare and a recognition of courts' and agencies limited capacities. As we discussed earlier, the rule of reason's focus on whether arrangements promote or suppress competition—often applied with a process rather than outcome orientation—is rationalized in part by the view that, although we care about the results of competition, it may often be easier for antitrust authorities to assess the process. Accordingly, we mentioned that courts often respond to arguments for exceptions—assertions that challenged behavior, despite being anticompetitive, is socially valuable—by stating that they should be addressed to the legislature (or, in some jurisdictions, to the enforcement agency that is authorized to promulgate exceptions).¹⁸⁴ In addition, the rule of reason is sometimes interpreted to imply more specialized rules, such as the *per se* prohibition on naked price-fixing. Likewise, for monopolization, it may readily be optimal to employ different, more specific tests for certain practices, wherein those tests as well as the general rule applicable when no such test has been developed are derived from broader guiding principles.¹⁸⁵

Competition policy is not seen as comprehensive regulation but as merely offering what may be viewed as rules of the competitive game. Antitrust law does not dictate players' specific moves, but certain types of behavior are prohibited. When such prohibitions can be stated in simple, general terms—like prohibitions on naked price-fixing—this strategy works well. When firms' behavior is more complex and subtle, often involving dynamic considerations—a common state

¹⁸² We offer an example below in subsection D.3 when discussing exclusive dealing.

¹⁸³ One also notes the inconsistency with the tendency to focus solely on consumer surplus in the horizontal merger context, as discussed in subsection IV.D.3.

¹⁸⁴ In the United States, where cases are often decided by lay juries who one supposes may find conflicting expert testimony to be confusing, there is added pressure to limit which cases can reach juries for fear of excessive false positives. Many court decisions in recent decades seem in part motivated by this consideration. Yet this approach raises problems because there is a single antitrust law that is equally applicable to decisions by the FTC, a specialized agency. Given that the FTC is independently authorized to enforce its own statute, courts could draw distinctions that give the FTC more flexibility, but they have tended not to do so. In other countries, where juries are not used and decisions in the first instance are typically made by competition authorities rather than general courts, the appropriate legal rules may well differ.

¹⁸⁵ And it indeed seems to be the case that rules on different exclusionary practices vary, as illustrated by our discussions of the legal tests for predatory pricing and exclusive dealing in subsections C.3 and D.3, below.

of affairs in monopolization cases—the task is more daunting. The challenge is especially great because, as we emphasized in our discussion of the monopoly power requirement, the law seeks to avoid excessive administrative costs, false positives, and perhaps most importantly the chilling of socially valuable business activity, objectives that are not easily achieved when the law is highly uncertain. Indeed, it is precisely the difficulty in defining and identifying exclusionary conduct that is seen as justifying the monopoly power screen that we examined in subsection 1.¹⁸⁶

C. Predatory Pricing

Predatory pricing is one of the most storied areas of antitrust law. Indeed, the Sherman Act resulted in no small part from concerns about predatory practices, and predatory practices were central to the 1911 *Standard Oil* case, which breathed life into §2.¹⁸⁷ After a century of debate, the antitrust treatment of predatory pricing still elicits strong reactions. Supporters of tough limitations on predatory pricing believe that they are necessary to prevent large, powerful firms from using their market positions and financial strength to deter entry and to drive existing smaller, weaker rivals from the market, thus fortifying their monopoly power. Skeptics argue with equal vigor that price cutting is the essence of competition, that imposing antitrust liability on a firm for setting its prices too low should only be done with great caution if ever, and that successful predation, if it happens at all, is extremely rare. Here we consider the relevant economic theory, empirical evidence, and appropriate legal test.

1. Economic Theory

McGee (1958) initiated the Chicago School attack on traditional concerns about predatory pricing by strongly challenging, as a factual matter, whether Standard Oil had in fact engaged in predatory pricing, as was commonly believed at the time. McGee also argued, as a theoretical matter, that predatory pricing would only be an optimal strategy under very stringent conditions that are rarely if ever met. First, he emphasized that it will typically be more profitable for a monopolist to acquire its rivals than to drive them out of business. However, this argument fails because such horizontal mergers would likely violate the antitrust laws (see Section IV) and, even if permitted, the monopolist might be able to acquire its smaller rivals on more favorable terms if it can establish a reputation as a predator (a possibility that McGee discounts). Second,

¹⁸⁶ The discussion in the present subsection has emphasized difficulties in defining the standard, whereas in practice difficulties in proving what actually happened are often more daunting. These problems will be addressed somewhat in the discussions to follow of predatory pricing and exclusive dealing. In attempting to differentiate efficient from anticompetitive behavior (however defined), most attention will be devoted to evidence bearing on the consequences of the practices under scrutiny. Another channel of proof, which is promising but also fraught with pitfalls, involves examining the defendant's internal decision making, which is sometimes done under the rubric of inquiries into intent. The promise is that many complex strategies with anticompetitive or efficiency-enhancing effects (or both) cannot be analyzed and implemented in a large firm without extensive communications that seem, in modern times, difficult to undertake without leaving paper and/or electronic trails. The pitfall is that, especially with decision-making by lay juries, aggressive rhetoric ("we will crush the competition") that is logically quite consistent with efficient behavior (for example, trimming costs, improving quality, marketing, and service) can, taken out of context, be mistaken for evidence of anticompetitive designs.

¹⁸⁷ *Standard Oil Co. of New Jersey v. United States*, 221 U.S. 1 (1911).

McGee pointed out that a monopolist might well lose more money than its smaller rivals during a period of predation because it is selling more units. This is an important point, but it does not eliminate the possibility of profitable predation, especially predation that involves price discrimination, with price cuts targeted at the prey's customers, or predation that establishes a reputation. Third, McGee noted that driving a rival from the market might do little to increase the predator's market power if the prey's productive assets remain intact and available to a new entrant. Lastly, McGee asked why the prey could not credibly survive the predation by drawing on internal funds or borrowing as needed.

Subsequently, there has arisen a burgeoning literature on the economics of predation, which has been surveyed by Ordover and Saloner (1989). We briefly note a few highlights. The early theory of predatory pricing was based on the superior financial resources of the incumbent monopolist in comparison with a smaller rival/entrant. Under this "deep pocket" theory, as formalized by Telser (1966) and later Benoit (1984), the smaller firm would earn positive profits, if not for the predation, but the smaller firm had a limited ability to sustain losses before it must exit the market. By depleting the rival's financial resources, or credibly threatening to do so, the predator can induce exit or deter entry. This theory assumes, however, that the prey cannot obtain the financial resources necessary to sustain itself despite the prospect of positive profits that await if it can survive the predation. This assumption would not be justified if a firm with substantial financial resources could enter the market in question or if the smaller firm could obtain a large line of credit to finance its operations. Either prospect would deter predation, and the financial resources would not in fact need to be drawn upon. Yet economic theorists have also explained how asymmetric information between potential lenders and the firm that seeks to borrow funds may interfere with this possibility. Bolton and Scharfstein (1990) show how deep-pocket predation can occur even if the prey and its lenders are sophisticated.

Much theoretical work on whether predation can be economically rational focuses on asymmetric information and uncertain time horizons in attempting to formalize the intuition that predators can develop a reputation that will not only drive the current prey from the market but deter others from entering. To frame the problem, consider the following simple game between a potential entrant and an incumbent firm. First, the entrant decides whether to incur the sunk costs necessary to enter the market. Next, the incumbent decides whether to accommodate entry or engage in predation. Suppose that the entrant will make positive profits if the incumbent accommodates entry but will lose money if the incumbent engages in predation. Following the Chicago School critique, suppose further that, viewing the problem as a one-shot game, predation is unprofitable—that is, the threat to predate is not credible—for whatever reason; perhaps the entrant will fight for a long period of time before exiting, or perhaps accommodation is quite profitable for the incumbent, so the opportunity cost of predation is high. The only subgame-perfect equilibrium is for the potential entrant to enter the market and for the incumbent to accommodate entry. Furthermore, if this game is repeated in a finite sequence of distinct markets involving the same incumbent, the unique subgame-perfect equilibrium is for the potential entrant in each of these markets to enter and for the incumbent to accommodate every time. Following Selten (1978), this (formerly) counter-intuitive backward-induction argument is known as the "chain-store paradox."

A body of subsequent work has shown that the chain-store paradox may well dissolve when any of a number of realistic dimensions is added. Milgrom and Roberts (1982, Appendix A) show

that the paradox is an artifact of the known, finite number of potential entrants: predation becomes credible so long as there always remains a sufficient probability that future potential entrants will arrive. And even with a finite number of periods, predation based on reputation is rather easily supported using game theory. Milgrom and Roberts (1982) and Kreps and Wilson (1982), in highly influential work, demonstrate the power of reputation and signaling to support credible predation. These papers rely heavily on the presence of asymmetric information, the essence of predation based on reputation being that the predator is signaling its willingness to engage in predation (for example, predatory behavior may signal low marginal cost). Scharfstein (1984) and Fudenberg and Tirole (1986) show how predation can also work by disrupting the ability of the entrant to determine whether remaining in the market will be profitable.

Collectively, these papers establish that predation can, in theory, be profitable for an established monopolist in a variety of plausible circumstances. This body of sophisticated theoretical work, however, cannot resolve the debate about whether predatory pricing is in fact a widespread threat to competition that must be met with vigorous antitrust enforcement or instead constitutes a phantom practice that rarely if ever occurs.

2. Empirical Evidence

The empirical evidence on predation has been hotly disputed since at least the 1950s. Regarding the landmark *Standard Oil* case, McGee (1958, p. 168) writes: “Judging from the Record, Standard Oil did not use predatory price discrimination to drive out competing refiners, nor did its pricing policies have that effect.” Ever since then, Chicago School proponents have complemented their theoretical attack on the economic logic underlying predatory pricing with the empirical claim that predatory pricing is either extremely rare or nonexistent.¹⁸⁸ In this tradition, Koller (1971) examines 26 cases, ranging from 1907 through 1965, in which he was able to obtain a substantial trial record and the defendant was found guilty of engaging in predatory pricing. By his count, only seven of these cases involved below-cost pricing with predatory intent; four of these seven cases involved predation to eliminate a rival, and three involved predation to acquire a rival or improve market discipline. Koller considers the predation to have been successful in only four cases.

The response to McGee and Koller began with Yamey (1972), who argued that predatory practices may not be nearly as rare as McGee suggests and provided an example of predation in the China-to-England ocean shipping business around 1890. Zerbe and Cooper (1982) update and expand on Koller’s study. Based on their examination of 40 predatory-pricing cases from 1940 through 1981, they recommend a modified version of the Areeda and Turner (1975) rule (see just below) to prevent predatory pricing, finding that it performs much better than a rule of *per se* legality. In this tradition, a significant empirical literature identifying instances of successful predation has emerged over the past twenty years. Burns (1986) presents evidence that predation by the tobacco trust enabled it to acquire its rivals—those who were targets of the

¹⁸⁸ As described by Baker (1994), the Chicago School view of predatory pricing is that it is akin either to a white tiger, an extremely rare creature, or to a unicorn, a complete myth. He calls theories of predation based on reputation effects an example of “Post-Chicago Economics,” which gives greater weight to market imperfections, such as those based on incomplete information, than does the Chicago School.

predation and others based on reputational effects—on more favorable terms. Ordover and Saloner’s (1989, p. 545) survey directly challenges McGee’s conclusions; citing Standard Oil documents, they state: “There is little doubt, however, that Standard Oil at least attempted to use pricing as a weapon to drive its rivals out.” Weiman and Levin (1994) argue that the Southern Bell company engaged in predation to protect and build its telephone monopoly. Morton (1997) finds related evidence of deep-pocket predation in merchant shipping. Genesove and Mullin (2006) find evidence of predatory pricing in the U.S. sugar refining industry before World War I. Bolton et al. (2000) assemble and discuss the body of empirical evidence of predatory pricing.¹⁸⁹

In the end, whether this evidence is sufficient to conclude that strong rules against predatory pricing are needed to protect competition is difficult to say. The rarity of predatory pricing convictions in the United States may simply indicate that the law is working well to deter this practice. On this score, it is interesting to note that many of the documented instances of predatory pricing are either prior to 1911, when the *Standard Oil* case put teeth into the Sherman Act, or from outside the United States

3. Legal Test

The contemporary discussion of rules to control predatory pricing can be dated to a highly influential paper by Areeda and Turner (1975). They expressed concern that the treatment of predatory pricing in the cases and in the literature did not clearly and correctly delineate which practices should be illegal and that fears of predatory pricing were overblown. Areeda and Turner were sharply critical of the case law, stating (p. 699): “Courts in predatory pricing cases have generally turned to such empty formulae as ‘below cost’ pricing, ruinous competition, or predatory intent in adjudicating liability. These standards provide little, if any, basis for analyzing the predatory pricing offense.” Areeda and Turner proposed a test for predation based on whether prices were below average variable cost. A cost-based test gradually won favor in the courts, most explicitly through the Supreme Court’s endorsement in *Brooke Group*, which insisted on “some measure of incremental cost” although without choosing a particular measure.¹⁹⁰ A cost-based test is also employed in the European Union, but the approach is less strict.¹⁹¹

¹⁸⁹ Isaac and Smith (1985) report results from a laboratory experiment designed to see if predatory pricing would arise. In their experiment, a subject controlling a large firm competed against another subject controlling a smaller firm. The firms produced with economies of scale, with the larger firm being more efficient and having superior financial resources. They also included sunk entry costs that would need to be incurred again in the event of exit. Despite these conditions, arguably favorable to predation, the subjects did not employ predatory pricing. On the other hand, Jung et al. (1994) find frequent predation in their experiment, in which a single monopolist plays a sequence of eight periods against a series of different entrants.

¹⁹⁰ *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 509 U.S. 209, 223 (1993). Our discussion in the text implicitly refers to monopolization claims under Sherman Act Section 2; claims are also possible under the Robinson-Patman Act, but *Brooke Group* greatly reduced the differences in the Acts’ requirements (moving the Robinson-Patman Act test closer to the Sherman Act monopolization test).

¹⁹¹ In addition to declaring prices below average variable cost by a dominant firm that eliminates competitors as abusive, prices above average variable cost but below average total cost that eliminate competitors might also be reached. *Tetra Pak International SA v. Commission*, Case C-333/94P [1996] ECR I-5951, ¶ 41; *ECS/Akzo II*,

A vigorous debate ensued over workable rules to control predatory pricing without generating a large number of false positives. Scherer (1976) criticized Areeda and Turner's analysis, Williamson (1977) and Baumol (1979) offered alternative tests, and further critical commentary was offered by Joskow and Klevorick (1979) and Ordover and Saloner (1989). More recently, there has been a further round of proposals and criticisms, including Bolton, Brodley, and Riordan (2000), Edlin (2002), and Elhauge (2003b). See also Areeda and Hovenkamp (2002, vol. 3).

To illuminate this controversy, it is useful to relate the question of the appropriate legal test for predatory pricing to two of our previous discussions. First, our survey of the economic theory in subsection 1, in explaining how predatory strategies could credibly deter entry or in some instances drive out rivals, made no reference to any cost-based test. The literature neither suggests that pricing below some particular concept of cost is a necessary condition nor that it is a sufficient one. (Indeed, many proposed alternatives are motivated by a belief that stringent cost-based tests are under-inclusive, that is, that they would exonerate much predatory behavior.) Also it should be kept in mind that the literature does not for the most part specifically indicate the effects of predatory pricing on consumer surplus or total welfare, which would be necessary to translate its results into a normative rule, but rather is focused on the circumstances in which predatory strategies are credible. In addition, the literature is not addressed to benign or beneficial price reductions in a variety of settings and thus does not indicate whether insisting that pricing be below a certain cost measure is a good way to avoid penalizing or chilling desirable behavior. Many believe that a cost-based test, perhaps one that uses average or marginal variable cost, would be a reasonable and administrable manner of identifying dangerous conduct while immunizing other conduct. It must be admitted, however, that this view reflects more a set of hunches than any precise combination of formal analysis and empirical evidence.

Second, different views on predation standards to a large extent track different perspectives on the proper antitrust standard toward exclusionary conduct generally (see subsection B.2), many of which were developed with predatory pricing in mind. For example, tests that condemn price cuts if and only if they are below marginal or variable cost are often defended because they reward productive efficiency, on the ground that in certain simple settings only less efficient rivals or entrants would be kept out of the market. This argument is consistent with defining exclusionary practices as those that reduce productive efficiency. As noted, however, a price reduction above this cost standard might still drive out slightly less efficient competitors whose presence might raise both consumer and total welfare. This sort of case motivates many of the proposed alternative predation rules, such as determining whether a defendant has made a short-run profit sacrifice or would not have engaged in a practice but for its effect of excluding rivals.

Another argument favoring a narrow prohibition on predatory pricing advanced by courts and commentators alike is that price cuts are a move in the right direction. This view does reflect a concern with consumer and total welfare. Of course, it is also a static view. Predatory pricing is problematic precisely because of a concern that higher prices will follow, a move in the wrong

Decision of the Commission, December 14, 1985, 1985 OJ L 374/1; Hilti, Decision of the Commission, December 22, 1987, 1988 OJ L 65/19.

direction. The important truth underlying this argument is that most price cutting in the economy is desirable, and thus it is a reminder that false positives and related chilling effects are particularly costly when contemplating punishment of low prices.

The overall balance between false positives and false negatives (and corresponding *ex ante* effects of each) depends in significant part on how one assesses the empirical evidence and the quality of the system of adjudication. If predatory pricing really is rare, as some Supreme Court pronouncements (based on partial and dated evidence) suggest, then the optimal test should reflect a disproportionate concern with false positives.¹⁹² The brief survey in subsection 2 does not support this view, but it remains the case that beneficial price cuts will vastly outnumber predatory ones, so heavy attention to false positives is nevertheless sensible. Of course, this is an important motivation for the monopoly power screen, as we discussed in Sections A and B.1.

Another important factor that we noted regarding exclusionary practices generally is the sophistication of decision-making. In the United States, predatory pricing is usually assessed by lay juries, who one might imagine would find conflicting expert testimony to be confusing and who might be sympathetic to a small firm driven out of business by a monopolist. Leading court opinions in the United States seem quite concerned about this matter and accordingly are reluctant to allow cases to proceed unless certain hurdles are overcome. Arguably, less caution is required if only the government (and not also disgruntled competitors) may initiate suit and if more expert agencies are responsible for applying tests for predation.

Additional complexity lurks beneath seemingly simple cost-based tests. Determining firms' costs in most settings is notoriously challenging, as we discussed in subsection II.D.1.a on the difficulty of measuring market power by observing the difference between price and marginal cost. One puzzle concerns the allocation of common costs. Even with a marginal cost test, common costs are often influenced at the margin, the only problem being that the extent of this phenomenon is difficult to measure. For example, how does one measure the cost of an additional employee? After including salary and fringes, one must think about secretaries and other support staff, rent and utilities on the space taken by the employee and others (keeping in mind that there are opportunity costs, as the space may well not otherwise remain vacant), and various central functions (time spent by the human resources department in searching for and hiring the employee, support from the computer department, and so forth). For nontrivial changes in output, which often are involved if a firm significantly lowers price in response to entry, these are just some of the complications in determining the cost of additional employees directly involved, say, in production, not to mention other costs. Note that the practice of ignoring myriad indirect costs tends systematically to underestimate marginal or variable cost

¹⁹² In *Brooke Group*, the Court merely recalled its remark in *Matsushita* on “the general implausibility of predatory pricing.” 509 U.S. at 227. *Matsushita*, a 1986 decision, in turn cited analytical arguments by legal commentators (but none of the literature mentioned in subsection 1 on modern economic analysis of predation) and the empirical papers by McGee and Koller but not that of others, such as Zerbe and Cooper, who study more cases. *Matsushita Electric Industrial Co., Ltd. v. Zenith Radio Corp.*, 475 U.S. 574, 588-90 (1986). As we have noted elsewhere, however, one should be generous in the assessment of court decisions when much of the relevant economic analysis and evidence may not have been presented to them by the litigants.

and by a sufficiently great magnitude to make tests requiring prices to be below those costs highly permissive in many settings.

Other factors cut in the opposite direction. For example, there may be learning by doing, in which case a proper analysis of current marginal costs suggests that a lower figure should be employed because account must be taken of how present output reduces the cost of future production. See Arrow (1962) and Spence (1981). Another difficulty concerns complementarity: selling more automobiles may increase future revenue from spare parts, increasing flights between destinations A and B may increase traffic on other routes to and from points A and B on the same airline, or added sales may improve familiarity with a brand shared by other products. None of these factors is easy to measure.

Yet another set of complications involves capacity and other forms of investment. In *American Airlines*, American placed additional flights on routes served by the entrant.¹⁹³ The appellate court focused on whether fares covered marginal costs on those flights. However, many of those passengers would otherwise have taken other American flights already serving the route, so true incremental revenue was less than appeared to be the case. Moreover, one supposes there was a substantial opportunity cost of diverting planes to the route in that, wherever they had previously been deployed, they probably contributed net revenue sufficient to justify American carrying the additional capacity.

Another possibility is that American might have maintained excess capacity precisely so that it could be deployed in response to new entry (and to the extent capacity is observable by prospective entrants, it would tend to discourage entry). More broadly, firms can invest in capacity to deter entry by making it easier to reduce price quickly, and they might make investments that lower future marginal costs, which may never be recovered directly but would allow them to charge lower prices in response to subsequent entry, which boosts credibility and also helps to avoid running afoul of predation tests based on variable costs. See Spence (1977) and Dixit (1979, 1980).

These latest examples begin to blur the distinction between predatory pricing and exclusionary practices more broadly. Ordover and Willig (1981) bridged this gap by offering examples of what they described as predatory product innovation, such as investments designed to make a product more attractive than a rival's product (while setting a price gap below the incremental value) or to make the products of a dominant incumbent firm incompatible with those of rivals or prospective entrants. Likewise, one can think of high expenditures on R&D designed to come up with a patented product that has low marginal cost. In all of these cases, ignoring prior investments gives a misleading picture of firms' possibly predatory behavior. On the other hand, subjecting complex investment, research, and product design decisions to intensive antitrust scrutiny may be quite dangerous. How is one to distinguish the firm that makes substantial expenditures in a new product or service that it hopes will have a huge ultimate payoff (think of Amazon.com, eBay, or Google) from the firm making a predatory investment? Hindsight is

¹⁹³ United States v. AMR Corp., 335 F.3d 1109 (10th Cir. 2003). For an analysis, see Edlin and Farrell (2004).

often twenty-twenty, but risky legitimate investments often fail, so it will usually be difficult to distinguish the cases even after the fact.

Returning to the law on predatory pricing proper, U.S. law (but not that in the European Union) has recently added an additional requirement, that a party alleging predation prove that it is likely that the defendant will ultimately recoup its interim losses.¹⁹⁴ On one hand, this requirement is certainly logical, for if recoupment is implausible, there will be no (or less than complete) ultimate loss of consumer or total welfare, and more importantly the inability to recoup casts doubt on whether predation has in fact occurred.¹⁹⁵ On the other hand, the requirement seems redundant. First, the monopoly power screen exists to distinguish cases in which anticompetitive conduct is plausible from those in which it is not. A firm would be unlikely to recoup its losses in situations in which there are close substitutes for its products or entry, supply substitution, and the like would impose significant constraints on price increases. But it is precisely these factors that negate the existence of monopoly power.¹⁹⁶ Second, if one is reasonably confident that predation has in fact occurred, that very fact gives rise to a logical inference of recoupment. After all, the firm would not sustain losses out of charity, so its own analysis suggests that recoupment is likely (more precisely, that on an expected basis, it will occur). It seems implausible that courts' or agencies' assessments in the context of litigation would be more accurate than those of the firm with its own funds on the line.

The main explanation for this seeming puzzle is the concern for false positives combined with uncertainty about the other legal elements for a monopolization claim. Doubts about the proof of predation are certainly understandable in light of the above discussion, but this uncertainty was the justification for a strong monopoly power screen, and it is harder to understand how the recoupment requirement supplements rather than repeats this screen. The central problem is to improve the ability to distinguish true predation from legitimate price-cutting. Yet alternative explanations for below-cost pricing—namely, promotion of new products (which can involve periods of sustained losses, which the above Internet examples illustrate)—also presuppose recoupment. That is, a firm will only be willing to suffer substantial losses in promoting an innovative product if the quasi-rents from subsequent above-marginal-cost pricing (discounted for time and probability) are greater. If there are close substitutes or ready prospects of entry (imitation), the costly campaign would not be undertaken.

In reflecting on the recoupment requirement, it is notable that it was announced in the *Brooke Group* case which involved an unfavorable setting in many respects. The case was brought by a competitor, which raises suspicions. The alleged recoupment was to be through oligopoly

¹⁹⁴ For U.S. law, the requirement is announced in *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 509 U.S. 209 (1993). For EU law, see *Tetra Pak International SA v. Commission*, Case C-333/94P [1996] ECR I-5951.

¹⁹⁵ The discussion in the text assumes that the pertinent cost test is one under which there are losses that need to be recouped.

¹⁹⁶ The degree to which this statement is true depends on how the monopoly power requirement is interpreted. If, as we suggest in subsection A.2 and throughout, it is interpreted with regard to the danger posed by the practice under consideration, then the recoupment requirement indeed seems fully redundant. Put another way, the recoupment requirement might be understood as a warning to assess monopoly power less in a vacuum and more in light of the challenged practice.

pricing, about which the Court expressed skepticism.¹⁹⁷ Finally, the finding of liability was made by a lay jury. In all, the Court's strong concern about false positives may have been warranted, but the logic of an independent recoupment requirement in addressing this concern remains unclear.

D. Exclusive Dealing

We now turn to exclusive dealing, an important form of non-price conduct that a monopolist might be able to use to fortify its dominant position. Exclusive dealing involves a supplier's conditioning its sales on its customers' refusal to purchase from its rivals.¹⁹⁸ In addition to being important in its own right, analysis of exclusive dealing illuminates the economics of a variety of other practices that we do not take up explicitly. We begin by considering possible anticompetitive effects, then briefly examine efficiencies, and finally discuss the legal test, illustrating the principles with a number of cases.

1. Anticompetitive Effects

We focus here on exclusive dealing imposed by an upstream monopolist, *M*, on downstream customers. In the scenario most common in antitrust cases, *M* imposes exclusivity on wholesalers or retailers (rather than on final consumers). Clearly, any customer who has agreed to deal exclusively with *M* cannot purchase from *M*'s rivals. In this mechanical sense, those rivals are excluded from selling to customers. But this would be formally true to some extent even without exclusivity because every unit bought from *M* is a unit that otherwise may have been purchased from *M*'s rival. To consider anticompetitive effects, it is necessary to introduce strategic considerations. Our analysis of these draws heavily on Whinston (2006, Chapter 4), a major contributor to the recent game-theoretic literature in this area, who provides a masterful treatment of exclusive dealing and other exclusionary vertical contracts.

Anticompetitive exclusion most plausibly arises when *M* requires its dealers to purchase only from itself, these dealers constitute a large proportion of the market, and profitable entry or continued survival requires the rival to achieve a scale greater than is possible if sales must be

¹⁹⁷ This skepticism may not have been warranted, for oligopoly pricing is hardly rare—and had previously been documented in the industry in *Brooke Group*, tobacco—and as we discussed in Section III punishments through price cuts are an important means of sustaining collusion.

¹⁹⁸ We generally will write in terms of explicit exclusive dealing but note in passing that a variety of seemingly distinct contractual arrangements, without explicit exclusivity, can have very similar economic effects. Consider, for example, a quantity discount in the form of a two-part tariff with a large fixed fee and a per-unit price that equals marginal cost. This pricing structure reduces the customer's incentive to purchase from alternative suppliers relative to the case in which the supplier charges a uniform price above marginal cost. Consider also discounts to buyers that purchase a large fraction of their needs from the incumbent supplier, such as was present in *LePage's, Inc. v. 3M*, 324 F.3d 141 (3d Cir. 2003). (In the limit, if the price is prohibitively high for a buyer who purchases less than 100% of its needs, the contract is economically equivalent to an exclusive dealing arrangement, but similar effects might be achieved far before this limit is reached.) A further variation on exclusive dealing arises when a buyer requires its suppliers not sell to its rivals. This was the fact pattern in *Toys "R" Us, Inc. v. Federal Trade Commission*, 221 F.3d 928 (7th Cir. 2000). Because the economic principles and analysis are similar, we only discuss the case in which it is the seller that imposes exclusivity on its customers.

limited to dealers not subject to exclusive-dealing contracts. A leading criticism of the possibility of anticompetitive exclusive dealing (and, as we shall discuss, of other allegedly anticompetitive contractual practices) comes from the Chicago School. In essence, the argument is that the dealers are harmed by anticompetitive exclusion because, if successful, the dealers will then be confronted by a monopoly; accordingly, it will be against their interests to enter into arrangements resulting in anticompetitive exclusion. As a corollary, if practices like exclusive dealing are nevertheless observed, it must be because they generate efficiencies rather than produce anticompetitive effects.¹⁹⁹ These views (and other criticisms of claims of anticompetitive exclusion) were initially launched in the 1950s by Director and Levi (1956) and were followed by a wave of related commentary, the most elaborate being Bork (1978).²⁰⁰

Over the course of this subsection, we will consider many aspects of this argument. We begin with the response that is probably most important in antitrust challenges to exclusive dealing: the presence of multiple (often very large numbers of) buyers, which leads to a free-rider problem in attempts to foil M 's anticompetitive design.²⁰¹ This point is elaborated informally in Kaplow (1985, pp. 531-36) and elsewhere, and it has been developed formally in subsequent work by Rasmusen et al. (1991), Innes and Sexton (1994), and Segal and Whinston (2000b).

To make this idea explicit, suppose as already hypothesized that a prospective entrant E must attract some critical mass of buyers to cover its fixed costs, and imagine that M attempts to enter into exclusive agreements with more than enough buyers so that E cannot profitably enter by serving those remaining.²⁰² None of these buyers would be pivotal, that is, none of them alone can induce E to enter by refraining from agreeing to exclusivity with M . Therefore, each would

¹⁹⁹ As noted by Farrell (2005), however, one could equally well conclude that exclusivity must generate some anticompetitive effects not captured in the simple model advanced by the Chicago School. Farrell (p. 468) characterizes the Chicago School argument on exclusive dealing as “a Rorschach test, and the inference often drawn from it is mere spin.”

²⁰⁰ For further references, discussion of objections in addition to the argument emphasized in the text, and critical commentary, see, for example, Kaplow (1985). To give some flavor of the Chicago School critique, consider Bork's (1978, pp. 306-07) remarks on *Standard Fashion Co. v. Magrane-Houston Co.*, 258 U.S. 346 (1922): “Standard can extract in the prices it charges retailers all that the uniqueness of its line is worth. It cannot charge the retailer that full worth in money and then charge it again in exclusivity the retailer does not wish to grant. To suppose that it can is to commit the error of double counting. . . . If Standard finds it worthwhile to purchase exclusivity from some retailers, the reason is not the barring of entry but some more sensible goal, such as obtaining the special selling effort of the outlet.” Bork goes on to say (p. 309): “A seller who wants exclusivity must give the buyer something for it. If he gives a lower price, the reason must be that the seller expects the arrangement to create efficiencies that justify the lower price. If he were to give the lower price simply to harm his rivals, he would be engaging in deliberate predation by price cutting, and that, as we have seen in Chapter 7, would be foolish and self-defeating behavior on his part.”

²⁰¹ Note that the Chicago School argument and the multiple buyers/free-rider response are applicable to a wide range of exclusionary practices, including predatory pricing (where the argument has also been raised, although less frequently).

²⁰² How many buyers E needs to serve depends on a number of factors, including M 's and E 's cost functions, the nature of demand for their products, and the mode of competitive interaction (in formal models, the extensive form of the game being studied). The basic argument in the text, however, depends only on this number being positive and not on how it is determined.

in fact agree to exclusivity in exchange for a very small additional payment. This argument supports an equilibrium in which all buyers agree to exclusivity, in exchange for an arbitrarily small transfer, and E is excluded from the market.²⁰³ Furthermore, this result follows even if E is more efficient than M . The key element that makes this equilibrium possible is a lack of coordination among the buyers: individual buyers, or even groups of buyers too small to offer the entrant enough business to enter profitably, cannot gain by refusing to sign exclusive contracts with M .²⁰⁴ But these contracts do harm all of the buyers and cause inefficiency. Enabling competition from E is a public good, and M can induce buyers to free ride, undermining that competition.

Subsequent work has explored the robustness of this exclusionary equilibrium. Whinston (2006) points out that this outcome might seem fragile since there arguably is another equilibrium in which none of the buyers agree to exclusivity. If M is constrained to make nondiscriminatory offers to the various buyers, and if the buyers can coordinate to the extent of selecting their Pareto-preferred equilibrium, none will agree to exclusivity in exchange for a *de minimis* payment. However, Segal and Whinston (2000b, 2003) show that exclusion is a robust outcome if M can make discriminatory offers to the various buyers. They also show that exclusion is easier to support if M makes sequential offers to the various buyers. As stated by Whinston (2006, p. 146): “More generally, the ability of the incumbent to approach buyers sequentially both reduces the cost of successful exclusion, and makes it more likely that the incumbent will find exclusion profitable. In fact, as the number of symmetric buyers grows large, so that each buyer becomes a very small part of aggregate demand, the incumbent is certain to be able to exclude for free.”²⁰⁵

Consider further how these ideas generalize to settings that often arise in practice in which exclusive dealing targets M 's existing rivals, not just potential entrants, and products are differentiated.²⁰⁶ Suppose now that E is an existing rival that, despite the presence of scale economies and its smaller relative size, is able to survive due to M 's high price and product differentiation, with E 's product especially well suited to some customers. In this scenario, M may successfully enter into exclusive arrangements with most dealers, thereby limiting E 's

²⁰³ Note that, since the required payment per buyer is trivial, it is essentially costless to M to sign up more buyers than necessary, insuring that no buyer will believe that there is any real possibility that it would be pivotal.

²⁰⁴ It is conceivable that a large group of buyers would attempt to agree not to deal with M , or at least not exclusively. In addition to the free-rider problem in organizing and enforcing such an agreement, it should be noted that it may well be illegal under the antitrust law. See subsection III.E.2. Typically, antitrust law does not allow buying cartels and other otherwise illegal arrangements to be justified on grounds that they create countervailing power.

²⁰⁵ Another counter-strategy would be for E to create or induce the entry of additional dealers or to bypass dealers and directly serve customers at the next level in the distribution chain. Often this will be infeasible or impose substantial costs on E ; think of products sold primarily through department stores, drug stores, or full-line wholesalers, where E supplies only one or a few of those products. This issue arose in the *Dentsply* case, discussed in subsection 3.

²⁰⁶ See also Salop and Scheffman (1983) and Krattenmaker and Salop (1986), who consider a variety of strategies designed to elevate rivals' costs, which may or may not induce exit, and the survey of this work in Ordovery and Saloner (1989, pp. 565-70).

ability to expand. Each of *M*'s customers would prefer to be free to deal with *E* but nevertheless may find it very costly to resist *M*'s exclusivity policy since this would require foregoing all purchases from *M*. In many settings, a wholesaler or retailer would find it difficult not to stock *M*'s products, which by assumption dominate the market. As a result, *E* may indeed conclude that making the necessary investments to expand its product line, manufacturing capacity, sales and distribution network, or advertising would not yield a sufficient return. *M*'s exclusivity may render unprofitable a strategy under which *E* gradually increases its share at the retail locations where *M* has traditionally been dominant. Many individual retailers choose to purchase from *M* on an exclusive basis, but collectively the retailers, and final consumers, are harmed in the long run by *M*'s exclusive dealing.²⁰⁷

Additional variations and qualifications should be mentioned. Most notably, one or a few large buyers may find it profitable to support entry. If there is a single buyer, the free-rider problem does not arise, and with only a few the problem is attenuated, especially if one is large enough to support an entrant—although note that its behavior may still convey a positive externality on others, so its incentives may not be sufficient. Fumagalli and Motta (2006) show, however, that if the buyers compete against each other, one or a few might grow large enough to support entry if they can obtain more favorable terms from *E* than their rivals did by signing exclusive contracts with *M*. Even if entry does not occur, the threat of large buyers to sponsor entry might induce *M* to offer them better deals, reducing *M*'s market power. On the other hand, as shown by Simpson and Wickelgren (2004), one buyer may have little incentive to resist anticompetitive exclusion of an upstream entrant so long as its rivals are equally disadvantaged; if all suffer similarly, higher input costs will largely be passed on to the next level, and it is those customers who will suffer from *M*'s continued monopoly. Yet another possibility, along the lines of Aghion and Bolton (1987), discussed further below, is that *M* might find it more profitable to sign contracts with stipulated damages that lead to entry by only the relatively more efficient potential entrants. Segal and Whinston (2000b) show that this outcome can arise in a model with multiple buyers.²⁰⁸ Finally, it should be noted that even when an incumbent monopolist can profitably exclude a rival using exclusive dealing, the welfare analysis is further complicated because some profitable entry is inefficient, as shown by Mankiw and Whinston (1986).²⁰⁹

²⁰⁷ As we will discuss below and illustrate with cases in subsection 3, formal exclusivity contracts are not necessary for this result. Similar effects may arise from pricing strategies (such as quantity or loyalty discounts) or threats (implicit or explicit) of reduced services or a complete cut-off if dealers also sell *E*'s wares.

²⁰⁸ Another interesting but complex strand of the literature examines situations in which there is direct competition between buyers to sign contracts that may be exclusive. Exclusive contracts can affect oligopolistic competition even if no firms are excluded from the market as a result. See Besanko and Perry (1993, 1994). Bernheim and Whinston (1998) study the effects of banning exclusive dealing in a number of distinct models, stating (p. 64): “We demonstrate that a ban may have surprisingly subtle and unintended effects.” For further discussion of the complexity and potential ambiguity of these models, see Whinston (2006, pp. 152-78). This literature overlaps with a broader literature on vertical integration and vertical contracting. See, for example, Hart and Tirole (1990), O'Brien and Shaffer (1992), McAfee and Schwartz (1994), and the survey by Rey and Tirole (2007).

²⁰⁹ The possible tradeoffs raise the question of the objectives of antitrust law and general principles governing exclusionary practices, discussed in subsection B.2.

Although the case of multiple buyers is central to most antitrust cases involving exclusive dealing, it is useful to consider more briefly the Chicago School argument about buyer resistance to exclusivity in the single-buyer case, one that has received substantial attention in the contract theory literature. Begin with a simple case in which the incumbent monopolist M has a constant marginal cost above that of the potential entrant E , who also bears a fixed cost to enter. In the case of interest, E enters if no exclusive dealing contract binds the sole buyer B to M , and we suppose that the price after entry is determined by Bertrand competition and thus equals M 's marginal cost.²¹⁰ Suppose further that M wishes, before E is on the scene, to bind B to buy from M even if E should enter, and also suppose that the proposed contract does not specify the price.²¹¹ In that case, B expects to pay the monopoly price, which is higher than what B would pay in the absence of the exclusive dealing contract. B 's loss in surplus from exclusivity is simply the sum of M 's gain in producer surplus and the deadweight loss, and this total is obviously less than M 's gain, by the amount of the deadweight loss. The most that M would pay for exclusivity is less than the least B would accept, so the Chicago School claim is valid under these assumptions. As pointed out by Whinston (2006, p. 139), this result does not rely on any specific bargaining model but rather reflects what he calls the bilateral bargaining principle: "if two parties (i) contract in isolation, (ii) have complete information about each others' payoffs, and (iii) lump-sum transfers are possible, then they will reach an agreement that maximizes their joint payoff." Readers will also recognize this claim as a version of the Coase theorem. Here, the joint payoff of M and B is reduced—by the amount of the deadweight loss—if they sign an exclusive contract. But this is not always the case, even with only a single buyer.²¹²

An important set of extensions allows for other types of contract between M and E . Aghion and Bolton (1987) examine an exclusive contract that stipulates the price P but allows B to breach and purchase instead from E (at whatever price E might offer) upon payment of damages G to M .²¹³ Aghion and Bolton show how M and B can select G to extract rents from E . Since B must pay G to M if it wishes to buy from E , E needs to offer B a better deal than if there were no exclusionary contract. It is in M and B 's joint interest to raise G just to the point that fully extracts E 's profits, sharing this gain between themselves. Note that this outcome is efficient and does not exclude E , which is by assumption more efficient than M .

However, Aghion and Bolton (1987) show that this relatively benign rent-shifting result does not generalize to situations in which there is uncertainty about E 's costs. In that case, perfect extraction from E is not possible since M and B do not know *ex ante* how much may be

²¹⁰ This will be the case so long as E 's profits are increasing in price at least up to that level. If not, price will be lower, but the conclusion to follow in the text will still hold.

²¹¹ This incompleteness may arise for the usual reasons, such as future uncertainty and problems of verifiability. If the contract did specify price—as allowed by Aghion and Bolton (1987), discussed below—then we might suppose that the price would equal M 's marginal cost with all rents extracted through an *ex ante* fixed charge.

²¹² In addition to the variations considered in the text to follow, Farrell (2005) points out that the conclusion does not even generalize to alternative models of post-entry duopoly between M and E . With Cournot duopoly, exclusion can be profitable and inefficient.

²¹³ Many of the ideas elaborated in this literature first appeared in Diamond and Maskin (1979). See also Chung (1992).

extracted. Setting G involves a familiar sort of tradeoff: raising G increases extraction from potential entrants with sufficiently low costs that they still enter but loses entry and thus forgoes extraction from those with higher costs. Thus, the privately optimal level of G will partially extract surplus from relatively efficient entrants and will exclude entrants that are not so efficient (but still more efficient than M). Through the contract, M and B act somewhat like a monopsonist purchasing from a distribution of potential entrants with different costs.²¹⁴ Even in this case, note that the buyer is not harmed by the use of stipulated damages on an *ex ante* basis. As emphasized by Farrell (2005), harm to the buyer from exclusivity is not possible in this simple setting because the buyer can always “just say no” by not agreeing to grant exclusivity to M . This result contrasts with that in the previously discussed models with multiple buyers.

The Chicago School argument about buyers’ resistance to exclusive dealing is an application of a broader critique applied to a wide range of exclusionary practices that is referred to as the “one-monopoly-profit theorem” and under related rubrics. The essence of the wider attack draws on the idea that monopolists, no matter how powerful, cannot get something for nothing. That is, the argument holds there is some level of profit or rent that inheres in a given monopoly position, and monopolists cannot extract other concessions (potentially anticompetitive ones or otherwise) without giving up something in return.

This general point and many of its applications served as a useful corrective to superficial arguments that used to be prominent in both court opinions and commentary. A monopolist cannot generate monopoly returns on related products or in other markets—in addition to the monopoly profits it already is earning through charging the monopoly price—simply by threatening to withhold the product on which it enjoys the monopoly. Monopoly prices do not rise without limit, as we elaborated in Section II. Rather, there is an optimal price characterized by the property that a slightly higher price loses as much profit due to lost sales as is gained by the heightened margin on retained sales. Buyers who would be lost are, by definition, at the margin. Therefore, demanding other concessions, if they impose any positive cost, will lose these buyers as well unless price is reduced or other countervailing inducements are offered.

Despite the important element of truth in this proposition, it is now well known that there are substantial qualifications. Indeed, the so-called “one-monopoly-profit theorem” literally holds only in very special cases not often thought to be realistic, such as when a monopolist ties its product to another that is used in fixed proportions and is available competitively.²¹⁵ Within a static framework, the extent of monopoly profits may depend on practices that facilitate price discrimination or that limit substitution, such as by tying the sales of a monopolized product to sales of partial substitutes at an appropriate margin when otherwise the substitutes can be

²¹⁴ Spier and Whinston (1995) point out that this line of argument requires that M and B be able to commit themselves to the terms of their initial contract; perfect and costless renegotiation between them after E enters would undermine their ability to extract rents from E . However, Spier and Whinston show that this sort of contract would nevertheless benefit M and B through its influence on incentives to undertake investments before E arrives on the scene.

²¹⁵ Many of these limitations have long been well known. Some are discussed and further references are offered in Kaplow (1985).

obtained at competitive prices. Many such practices have indeterminate effects on consumer and total welfare.

In a dynamic framework—especially when one also introduces externalities (which may be common with regard to effects on competition), asymmetric information, and other strategic dimensions—there are many more possibilities, including ones with anticompetitive effects. Moreover, most claims about exclusionary practices are expressly of a dynamic character. (Consider, for example, many proposed tests for exclusionary practices that we discussed in subsection B.2, such as whether a short-run sacrifice is involved or whether a practice would be profitable but for its ultimate effect of excluding rivals.) The point that analysis can change qualitatively in such dynamic settings is well illustrated by the foregoing discussion of exclusive dealing with multiple buyers and also by our examination of credible predation in subsection C.1. Indeed, substantial bodies of literature in industrial organization over the past few decades have been devoted to settings in which the one-monopoly-profit theorem does not hold. In any event, we have seen that it does not negate the anticompetitive potential of exclusive dealing.

2. Efficiencies

That exclusive arrangements can promote efficiency may be inferred from their use in situations where meaningful market power is clearly absent, such as in many employment contracts. Employees or members of a partnership may be forbidden from working elsewhere in order to avoid diversion of effort and to limit their ability to take personal advantage of opportunities developed by the enterprise. Marvel (1982) developed these notions in a context more pertinent to exclusive dealing arrangements challenged under the antitrust laws. For example, a manufacturer that makes investments to attract customers to a retailer may be concerned that the retailer would free ride by diverting these customers to competitors' products if not precluded from doing so by some form of exclusivity. As another example, Masten and Snyder (1993) revisit the famous *United Shoe* case, arguing that the contractual provisions inducing shoe manufacturers to exclusively use United Shoe's machines protected the investments made by United Shoe in training shoe manufacturers to organize their production processes more efficiently.²¹⁶

Subsequent literature on contract theory has refined our understanding of the underlying mechanism. In Segal and Whinston's (2000a) model, a buyer and seller, subsequent to entering into a contract, independently make noncontractible investments, after which they bargain over the terms of trade. If the initial contract is not exclusive, the buyer has the option of turning to an alternative supplier of the product. In a model in which the buyer only needs one unit and the seller can make investments that are specific to the relationship with the buyer—that is, they provide no value to either buyer or seller if the two do not end up dealing with each other—Segal and Whinston show that exclusivity has no effect on the seller's investment incentives. Exclusivity reduces the buyer's threat point and thus raises the seller's *ex post* payoff, but in a way that is unaffected by the seller's investment.

²¹⁶ *United States v. United Shoe Machinery Corp.*, 110 F. Supp. 295 (D. Mass. 1953), affirmed per curiam, 347 U.S. 521 (1954).

This result suggests that pro-competitive justifications for exclusivity based on free riding and investment incentives require investments that are not entirely relationship specific. Segal and Whinston (2000a) show that exclusivity promotes seller investments that are also valuable to the buyer when dealing with third parties, but discourages seller investments that raise value to the buyer from remaining with the seller relative to switching to third parties. Opposite results apply for buyer investments. The previously noted examples of seemingly efficient exclusive dealing fit this pattern in that they involve seller investments that the buyer can exploit in dealing with alternative suppliers. Although Segal and Whinston's analysis solidifies our understanding, they also point out that the full welfare analysis is more complex because increased investment need not mean increased welfare.

3. Legal Test

It has long been believed that exclusive dealing contracts and related arrangements have the potential both to be anticompetitive and to promote efficiency. Accordingly, U.S. law has applied a balancing test along the lines of that under the rule of reason, which we discussed in subsection III.E.2.²¹⁷ As a further legal note, contractual exclusivity may be challenged under a variety of U.S. antitrust provisions. In addition to monopolization under Sherman Act Section 2, it may be reached under Section 1 since an agreement is involved, under Clayton Act Section 3's prohibition on contract provisions that anticompetitively restrict dealings with competitors, and under FTC Act Section 5's broad proscription against unfair methods of competition. The same is true of a variety of other forms of vertical restraints, including tying. As we noted in subsection IV.D.1 in connection with horizontal mergers, however, there has been a growing convergence in treatment regardless of the particular statutory provision invoked.

To state that the law applies a balancing test or, more particularly, applies a rule of reason, does not convey a very clear sense of how it actually operates, so it is useful to consider some cases. The *Standard Stations* decision in 1949 reflects the much stricter attitude of the Supreme Court at that time period.²¹⁸ Standard Oil had exclusive supply contracts with 16% of the retail outlets in the geographic market, most of which were terminable at 6 month intervals upon giving 30 days notice. Although this arrangement does not seem to constitute an insuperable barrier to an entrant or a rival seeking to expand (despite the fact that other suppliers also used similar arrangements), the Court affirmed a determination that it was anticompetitive. A stronger case was presented in *Lorain Journal*, where an incumbent newspaper with a local news and

²¹⁷ *Standard Oil Co. of California v. United States*, 337 U.S. 293 (1949), discussed below, recognized that exclusive arrangements were potentially efficient even though it judged Standard Oil's arrangements harshly. Subsequent cases, including some of the others discussed below, have required a greater demonstration of anticompetitive effects and also have more clearly acknowledged that efficiencies count, even if they were unconvinced by those presented by the defendants. See Hovenkamp (1998, vol. 11, ch. 18D). In the European Union, Article 82 on abuse of a dominant position encompasses exclusivity and related agreements (such as loyalty or fidelity rebates and various quantity discounts) imposed by a dominant supplier. Also covered are "English clauses" (which in the United States are usually called price-matching clauses) that allow purchasers to buy from rivals offering lower prices, but only if they first inform the contract supplier and that supplier is unwilling to make the sale at an equivalent price.

²¹⁸ *Standard Oil Co. of California v. United States*, 337 U.S. 293 (1949).

advertising monopoly (in 1948) was found guilty of attempted monopolization for refusing to carry ads of those who also advertised on the newly entered radio station.²¹⁹

More recently, variations on the exclusivity theme have appeared in many phases of the litigation involving Microsoft. In the mid-1990s, the government challenged and Microsoft ultimately agreed to cease the use of per-processor licensing fees for its operating system. Computer manufacturers who had wished to load Microsoft's operating system on some of their computers were charged for loading it on all of the computers they shipped, as a condition for dealing with Microsoft. Although not literally barred from dealing with competitors, computer manufacturers were discouraged from doing so since they had to pay for Microsoft's operating system even on computers shipped with an alternative operating system (or with none). Subsequent litigation successfully challenged other features of Microsoft's contracting and operating system design that exhibited some exclusivity.²²⁰ In another recent case, *Dentsply*, the leading supplier of artificial teeth with a 75-80% market share was found to have violated §2 for imposing exclusivity on its dealers.²²¹

A number of features of these cases are notable. First, all involved multiple buyers, although two of the dealers in *Dentsply* did have substantial market shares. Also, except for *Standard Stations*, the defendants seemed to possess monopoly power, and successful entry and ultimate expansion would seem to have required significant scale.²²² These features are consistent with the analysis of anticompetitive effects presented in subsection 1. Regarding efficiencies, none seemed apparent in *Lorain Journal*, and efficiency justifications offered by Microsoft and *Dentsply* were found to be unconvincing.

Like other allegedly exclusionary practices, exclusive dealing presents potentially difficult problems of balancing, raising both factual questions and issues about the precise content of the

²¹⁹ *Lorain Journal Co. v. United States*, 342 U.S. 143 (1951). This case is often discussed along with *United States v. Griffith*, 334 U.S. 100 (1948), in which the Supreme Court found monopolization where a chain of movie theatres with monopolies in many towns insisted on certain exclusive rights in all towns. Questions on both cases that outline the analysis in subsection 1 for the situation involving multiple buyers (though, for *Griffith*, it was multiple suppliers) appear in Areeda, Kaplow, and Edlin (2004, ch. 3).

²²⁰ *United States v. Microsoft Corp.*, 253 F.3d 34 (D.C. Cir. 2001).

²²¹ *United States v. Dentsply International, Inc.*, 399 F.3d 181 (3d Cir. 2005). An interesting feature of this case is that *Dentsply* did not formally have exclusive contracts with its dealers, but it did have supply arrangements that were terminable by it at will, combined with a formal policy of terminating dealers who carried competing products (subject to some grandfathered exceptions). The government convinced the court, based in large part on a series of actual events, including threatened terminations followed by dealers' discontinuance of competitors' products, that dealers did not believe they could be successful without *Dentsply* products. This example illustrates that exclusive dealing policies can have anticompetitive effects even without the use of formal exclusive-dealing contracts, much less long-term exclusive-dealing contracts.

²²² In *Dentsply*, there had long been a number of small suppliers, but the court was convinced that, without access to most dealers (some had access to certain dealers and many attempted to sell directly to the next level in the distribution chain, dental laboratories), the rivals could not realistically expand.

legal test, the subject of subsection B.2.²²³ A further challenge is raised by the possibility (as in horizontal merger cases) that both anticompetitive effects and efficiencies may be present simultaneously, given that the logics underlying the two considerations are essentially independent. Accordingly, it is also possible that a highly anticompetitive exclusive arrangement would involve no short-term profit sacrifice by the monopolist. As explained in subsection 1, with large numbers of buyers, exclusivity that has no efficiency consequences might be secured at a trivial cost; hence, even the slightest efficiency benefit would produce immediate (even if modest) gains, along with more substantial future profits due to anticompetitive effects. Note that in such cases any issue of recoupment would likewise be moot.

One way that the law addresses the problem of distinguishing legitimate and harmful exclusive dealing is through the monopoly power screen, which readily filters out a vast proportion of exclusive arrangements that may be efficient. Think of routine employment contracts, exclusivity provisions in partnerships, and most of the countless products sold (typically but not always without exclusivity provisions, in fact) in department stores and drug stores and distributed by wholesalers of all sorts. Relatedly, in subsection A.2 we discussed how, in addition to the traditional monopoly power requirement, one can also assess whether a challenged practice has any prospect of significantly damaging competition by stipulating that it has the alleged anticompetitive effect and then determining how large that effect could be. In addressing efficiencies, courts' approach has a feature analogous to the requirement in horizontal merger cases that efficiencies be merger specific (see subsection IV.D.3). Thus, for purported benefits of exclusivity and other restrictive contractual features, courts typically ask whether there exists a less restrictive alternative. For example, if the manufacturer provides some service or training, it might impose separate charges rather than employ exclusivity to prevent free riding.²²⁴

VI. Conclusion

Having surveyed several key economic underpinnings of antitrust policy and applied the lessons to the core features of existing regimes, we have seen that economics has had a tremendous influence on the law, but also that there is still much unfinished business for economists and lawyers alike. For each of our four main topics, we have noted that modern antitrust law in the United States—and to a substantial degree in the European Union—is aimed in large part at economic objectives and heavily employs economic tools in achieving them. At the same time,

²²³ Regarding the facts, Whinston (2006, pp. 189-97) indicates that there is “remarkably limited” empirical evidence on the motives and effects of exclusive contracting. Of course, it still may be possible to make determinations in specific cases. The court in *Dentsply* summarily dismissed the efficiency claim as “pretextual,” seeing it to be clearly contradicted by the evidence. In *Microsoft*, evaluation of Microsoft’s efficiency claims was more difficult regarding at least some of its challenged practices, particularly those in which the *de facto* exclusivity was not a feature of contracts but of the product (operating system) itself.

²²⁴ This approach is no panacea, however. A frequent problem with such alternatives—paralleling a problem with conduct remedies in monopolization cases—is that the terms, including the price, may need to be monitored. However, if exclusivity were forbidden, the monopolist that has effective services to offer would not have an incentive to charge a prohibitive price.

in every field there appear to be notable divergences, and ones that cannot fully be explained by administrative convenience or limitations on institutional competence. Some discrepancies may be the product of conscious choice; others no doubt reflect the inevitable lag in the dissemination of economic principles.

At least as important for economists, our primary audience, are the many ways that existing theoretical work and empirical methods, valuable as they are, do not yet adequately address many of the questions that those who formulate and apply competition policy need to answer. For example, we noted that antitrust law on collusion seems quite interested in forms of communication, whereas this matter plays a relatively minor role in economic models and empirical work. In many instances, the problem may be that the legal regime does not ask the right questions. Even in such cases, however, answers would be helpful, if for no other reason than to see in what sense and to what extent antitrust decisions have been led astray. In other cases, one cannot expect the law to answer economic questions sensibly, in real time, when leading economic research has not yet done so. One point of particular interest in many areas of competition policy is that the law is very concerned with minimizing error, especially false positives and associated chilling effects. But it is difficult to calibrate legal tests without better-informed priors. Forming such priors requires, in turn, an empirical and theoretical understanding of many legitimate practices as well as of the anticompetitive ones that naturally are the focus of legal disputes. This knowledge is important because, in many actual cases that must be adjudicated, these practices are not easily distinguished from each other.

VII. References

- Abreu, D. (1986), "Extremal equilibria of oligopolistic supergames", *Journal of Economic Theory* 39:191-225.
- Abreu, D., D. Pearce and E. Stacchetti (1986), "Optimal cartel equilibria with imperfect monitoring", *Journal of Economic Theory* 39:251-269.
- Abreu, D., D. Pearce and E. Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring", *Econometrica* 58:1041-1063.
- Aghion, P. and P. Bolton (1987), "Contracts as a barrier to entry", *American Economic Review* 77:388-401.
- American Bar Association, Section of Antitrust Law (2005), *Market Power Handbook* (American Bar Association, Chicago).
- Andrade, G., M. Mitchell and E. Stafford (2001), "New evidence and perspectives on mergers", *Journal of Economic Perspectives* 15(2):103-120.
- Areeda, P. and H. Hovenkamp (2002), *Antitrust Law: An Analysis of Antitrust Principles and Their Application*, vols. 3, 3A, 6, 2nd ed. (Aspen Law & Business, New York).

- Areeda, P. and D. Turner (1975), "Predatory pricing and related practices under Section 2 of the Sherman Act", *Harvard Law Review* 88:697-733.
- Areeda, P., H. Hovenkamp and J. Solow (1998), *Antitrust Law: An Analysis of Antitrust Principles and Their Application*, vols. 4 and vol. 4A, revised ed. (Aspen Law & Business, New York).
- Areeda, P., L. Kaplow and A. Edlin (2004), *Antitrust Analysis*, 6th ed. (Aspen Publishers, New York).
- Armstrong, M. and R. Porter, eds. (2007), *Handbook of Industrial Organization*, vol. 3 (Elsevier Science Publishers, New York).
- Arrow, K. (1962), "The economic implications of learning by doing", *Review of Economic Studies* 29:155-173.
- Athey, S. and K. Bagwell (2001), "Optimal collusion with private information", *RAND Journal of Economics* 32:428-465.
- Athey, S. and K. Bagwell (2004), "Collusion with persistent cost shocks," working paper, Harvard University.
- Athey, S., K. Bagwell and C. Sanchirico (2004), "Collusion and price rigidity", *Review of Economic Studies* 71:317-349.
- Bagwell, K. and R. Staiger (1997), "Collusion over the business cycle", *RAND Journal of Economics* 28:82-106.
- Bajari, P. and G. Summers (2002), "Detecting collusion in procurement auctions", *Antitrust Law Journal* 70:143-170.
- Baker, J. (1994), "Predatory pricing after *Brooke Group*: An economic perspective", *Antitrust Law Journal* 62:585-603.
- Baker, J. (1999), "Econometric analysis in *FTC v. Staples*", *Journal of Public Policy & Marketing* 18:11-21.
- Baker, J. (2002), "Mavericks, mergers and exclusion: Proving coordinated competitive effects under the antitrust laws", *New York University Law Review* 77:135-203.
- Baker, J. (2003a), "The case for antitrust enforcement", *Journal of Economic Perspectives* 17(2):27-50.
- Baker, J. (2003b), "Responding to developments in economics and the courts: Entry in the merger guidelines", *Antitrust Law Journal* 71:189-206.
- Baker, J. and T. Bresnahan (1985), "The gains from merger or collusion in product-differentiated industries", *Journal of Industrial Economics* 33:427-444.

- Baker, J. and T. Bresnahan (1988), "Estimating the residual demand curve facing a single firm", *International Journal of Industrial Organization* 6:283-300.
- Baker, J. and D. Rubinfeld (1999), "Empirical methods in antitrust litigation: Review and critique", *American Law and Economics Review* 1:386-435.
- Barton, D. and R. Sherman, (1984), "The price and profit effects of horizontal merger: A case study", *Journal of Industrial Economics* 33:165-177.
- Baumol, W. (1979), "Quasi-permanence of price reductions: A policy for prevention of predatory pricing", *Yale Law Journal* 89:1-26.
- Bellamy, C. and G. Child, (2001), *European Community Law of Competition*, 5th ed. (Sweet & Maxwell, London).
- Benoit, J.-P. (1984), "Financially constrained entry in a game with incomplete information", *RAND Journal of Economics* 15:490-499.
- Benoit, J.-P. and V. Krishna (1987), "Dynamic duopoly: Prices and quantities", *Review of Economic Studies* 54:23-35.
- Bernheim, B.D. and D. Ray (1989) "Collective dynamic consistency in repeated games", *Games and Economic Behavior* 1:295-326.
- Bernheim, B.D. and M. Whinston (1990), "Multimarket contact and collusive behavior", *RAND Journal of Economics* 21:1-26.
- Bernheim, B.D. and M. Whinston (1998), "Exclusive dealing", *Journal of Political Economy* 106:64-103.
- Berry, S. (1994), "Estimating discrete-choice models of product differentiation", *RAND Journal of Economics* 25:242-262.
- Berry, S. and P. Reiss (2007), "Empirical models of entry and market structure," in: M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, vol. 3 (Elsevier Science Publishers, New York), forthcoming, available at <http://www.stanford.edu/~preiss/hand-entry.pdf>.
- Berry, S., J. Levinsohn and A. Pakes (1995), "Automobile prices in market equilibrium", *Econometrica* 63:841-890.
- Besanko, D. and M. Perry (1993), "Exclusive dealing in a spatial model of retail differentiation", *International Journal of Industrial Organization* 12:297-329.
- Besanko, D. and M. Perry (1994), "Equilibrium incentives for exclusive dealing in a differentiated products oligopoly", *RAND Journal of Economics* 24:646-667.

- Bolton, P. and D. Scharfstein (1990), “A theory of predation based on agency problems in financing contracting”, *American Economic Review* 80:93-106.
- Bolton, P., J. Brodley and M. Riordan (2000), “Predatory pricing: Strategic theory and legal policy”, *Georgetown Law Journal* 88:2239-2330.
- Borenstein, S. (1990), “Airline mergers, airport dominance, and market power”, *American Economic Review Papers and Proceedings* 80(2):400-404.
- Borenstein, S. (2004), “Rapid price communication and coordination: The airline tariff publishing case (1994)”, in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 233-251.
- Borenstein, S. and A. Shepard (1996), “Dynamic pricing in retail gasoline markets”, *RAND Journal of Economics* 27:429-451.
- Bork, R. (1978), *The Antitrust Paradox: A Policy at War with Itself* (Basic Books, New York).
- Bresnahan, T. (1989), “Empirical studies of industries with market power,” in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 2 (North-Holland, Amsterdam) 1011-1057.
- Bresnahan, T. and S. Salop (1986), “Quantifying the competitive effects of production joint ventures”, *International Journal of Industrial Organization* 4:155-175.
- Brock, W. and J. Scheinkman (1985), “Price setting supergames with capacity constraints”, *Review of Economic Studies* 52:371-382.
- Bulow, J. and P. Pfleiderer (1983), “A note on the effect of cost changes on prices”, *Journal of Political Economy* 91:182-185.
- Bulow, J. and C. Shapiro (2004), “The BP Amoco–ARCO Merger: Alaskan crude oil (2000)”, in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 128-149.
- Burns, M. (1986), “Predatory pricing and the acquisition cost of competitors”, *Journal of Political Economy* 94:266-296.
- Capps, C., D. Dranove, S. Greenstein and M. Satterthwaite (2002), “Antitrust policy and hospital mergers: Recommendations for a new approach”, *Antitrust Bulletin* 27:677-714.
- Carlton, D. (2004), “Why barriers to entry are barriers to understanding”, *American Economic Review Papers and Proceedings* 94(2):466-470.
- Carlton, D. and J.M. Perloff (2005), *Modern Industrial Organization*, 4th ed. (Addison Wesley, Boston).

- Chamberlin, E. (1933), *The Theory of Monopolistic Competition* (Harvard University Press, Cambridge).
- Chung, T.-Y. (1992), "On the social optimality of liquidated damage clauses: An economic analysis", *Journal of Law, Economics, & Organization* 8:280-305.
- Coase, R. (1937), "The nature of the firm", *Economica* 4:386-405.
- Coate, M. and S. Ulrick (2005), "Transparency at the Federal Trade Commission: The horizontal merger review process 1996-2003", Bureau of Economics, Federal Trade Commission, available at www.ftc.gov/os/2005/02/0502economicissues.pdf.
- Compte, O. (1998), "Communication in repeated games with private monitoring", *Econometrica* 66:597-626.
- Compte, O., F. Jenny and P. Rey (2002), "Capacity constraints, mergers and collusion", *European Economic Review* 46(1):1-29.
- Connor, J. (2004), "Global cartels redux: The Amino Acid Lysine Antitrust Litigation (1996)", in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 252-276.
- Connor, J.M. (2007), "Price-fixing overcharges: Legal and economic evidence", in: J.B. Kirkwood, ed., *Research in Law and Economics*, vol. 22 (JAI/Elsevier, Amsterdam), forthcoming.
- Cooper, R., D. DeJong, R. Forsythe and T. Ross (1989), "Communication in the battle of the sexes game: Some experimental results", *RAND Journal of Economics* 20:568-587.
- Cooper, T. (1986), "Most-favored-customer pricing and tacit collusion", *RAND Journal of Economics* 17:377-388.
- Cournot, A.A. (1838), *Researches into the Mathematical Principles of the Theory of Wealth*, English ed. (Kelley, New York).
- Cramton, P. and J. Schwartz (2000), "Collusive bidding: Lessons from the FCC spectrum auctions", *Journal of Regulatory Economics* 17(4):229-252.
- Crandall, R. and C. Whinston (2003), "Does antitrust policy improve consumer welfare: Assessing the evidence", *Journal of Economic Perspectives* 17:3-26.
- Crawford, V. (1998), "A survey of experiments on communication via cheap talk", *Journal of Economic Theory* 78:286-298.
- Dabbah, M. (2004), *EC and UK Competition Law* (Cambridge University Press, Cambridge).
- Dalkir, S. and F. Warren-Boulton (2004), "Prices, market definition, and the effects of merger: Staples–Office Depot (1997)", in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust*

- Revolution: Economics, Competition, and Policy, 4th ed. (Oxford University Press, New York) 52-72.
- Davidson, C. and R. Deneckere (1984), "Horizontal mergers and collusive behavior", *International Journal of Industrial Organization* 2:117-132.
- Davidson, C. and R. Deneckere (1990), "Excess capacity and collusion", *International Economic Review* 31:521-541.
- Demsetz, H. (1973), "Industry structure, market rivalry, and public policy", *Journal of Law and Economics* 16:1-10.
- Deneckere, R. and C. Davidson (1985), "Incentives to form coalitions with Bertrand competition", *RAND Journal of Economics* 16:473-486.
- Department of Justice, Antitrust Division, (2006), *Workload Statistics, FY 1996-2005*, <http://www.usdoj.gov/atr/public/workstats.htm>.
- Diamond, P. and E. Maskin (1979), "An equilibrium analysis of search and breach of contract, I: Steady states", *Bell Journal of Economics* 10:282-316.
- Director, A. and E. Levi (1956), "Law and the future: Trade regulation", *Northwestern University Law Review* 51: 281-296.
- Dixit, A. (1979), "A model of duopoly suggesting a theory of entry barriers", *Bell Journal of Economics* 10: 20-32.
- Dixit, A. (1980), "The role of investment in entry-deterrence", *Economic Journal* 90:95-106.
- Edlin, A. (1997), "Do guaranteed-low-price policies guarantee high prices, and can antitrust rise to the challenge?", *Harvard Law Review* 111:528-575.
- Edlin, A. (2002), "Stopping above-cost predatory pricing", *Yale Law Journal* 111:941-991.
- Edlin, A. and E. Emch (1999), "The welfare losses from price matching policies", *Journal of Industrial Economics* 47:145-167.
- Edlin, A. and J. Farrell (2004), "The American Airlines case: A chance to clarify predation policy (2001)", in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 502-527.
- Elhauge, E. (2003a), "Defining better monopolization standards", *Stanford Law Review* 56:253-344.
- Elhauge, E. (2003b), "Why above-cost price cuts to drive out entrants are not predatory and the implications for defining market power", *Yale Law Journal* 112:681-828.

- Ellison, G. (1994), "Theories of cartel stability and the joint executive committee", *RAND Journal of Economics* 25:37-57.
- Elzinga, K. and T. Hogarty (1973), "The problem of geographical market delineation in antimerger suits", *Antitrust Bulletin* 18:45-81.
- Elzinga, K. and D. Mills (2004), "The brand name prescription drugs antitrust litigation (1999)", in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 301-320.
- Epstein, R. and D. Rubinfeld (2001), "Merger simulation: A simplified approach with new applications", *Antitrust Law Journal* 69:883-919.
- Epstein, R. and D. Rubinfeld (2004), "Merger simulation with brand-level margin data: Extending PCAIDS with nests", *Advances in Economic Analysis and Policy* 4:1-26.
- European Union (2001), "Guidelines on the applicability of Article 81 of the EC Treaty to horizontal cooperative agreements", European Commission, *Official Journal of the European Communities*, 6.1.2001, C 3/2.
- European Union (2004a), "Guidelines on the assessment of horizontal mergers", European Commission, Regulation 139/2004, 1 May.
- European Union (2004b), "Guidelines on the assessment of horizontal mergers under the Council regulation on the control of concentrations between undertakings", *Official Journal of the European Union*, February, 2004/C 31/03.
- Evans, W.N. and I.N. Kessides (1994), "Living by the 'Golden Rule': Multimarket contact in the U.S. airline industry", *Quarterly Journal of Economics* 109:341-366.
- Farrell, J. (2000), "Renegotiation in repeated oligopoly interaction", in: G. Myles and P. Hammond, eds., *Incentives, Organisation, and Public Economics: Papers in Honour of Sir James Mirrlees* (Oxford University Press, New York) 303-322.
- Farrell, J. (2005), "Deconstructing Chicago on exclusive dealing", *Antitrust Bulletin* 50:465-480.
- Farrell, J. and M.L. Katz (2006), "The economics of welfare standards in antitrust", *Competition Policy International* 2:3-28.
- Farrell, J. and E. Maskin (1989) "Renegotiation in repeated games", *Games and Economic Behavior* 1:327-360.
- Farrell, J. and M. Rabin (1996), "Cheap talk", *Journal of Economic Perspectives* 10(3):103-118.
- Farrell, J. and C. Shapiro (1990a), "Asset ownership and market structure in oligopoly", *RAND Journal of Economics* 21:275-292.

- Farrell, J. and C. Shapiro (1990b), "Horizontal mergers: An equilibrium analysis", *American Economic Review* 80:107-126.
- Farrell, J. and C. Shapiro (2001), "Scale economies and synergies in horizontal merger analysis", *Antitrust Law Journal* 68:685-710.
- Federal Trade Commission (1999), "A study of the Commission's divestiture process", Bureau of Competition, available at <http://www.ftc.gov/os/1999/08/divestiture.pdf>.
- Federal Trade Commission (2002), "Understanding mergers: Strategy and planning, implementation and outcomes", Bureau of Economics, Merger Roundtable Papers, available at <http://www.ftc.gov/be/rt/mergerroundtable.htm>.
- Federal Trade Commission and U.S. Department of Justice (1992), "Horizontal merger guidelines" (amended in 1997), at <http://www.usdoj.gov/atr/public/guidelines/hmg.htm>.
- Federal Trade Commission and U.S. Department of Justice (2003), "Merger challenges data: Fiscal years 1999-2003", available at <http://www.ftc.gov/os/2003/12/mdp.pdf>.
- Federal Trade Commission and U.S. Department of Justice (2004), "Joint workshop on merger enforcement", February 17-19, <http://www.ftc.gov/bc/mergerenforce/index.html>.
- Federal Trade Commission and U.S. Department of Justice (2006), "Commentary on the horizontal merger guidelines", available at www.ftc.gov/os/2006/03/CommentaryontheHorizontalMergerGuidelinesMarch2006.pdf.
- Fisher, F. and J. McGowan (1983), "On the misuse of accounting rates of return to infer monopoly profits", *American Economic Review* 73:82-97.
- Frech, H.E., J. Langenfeld and R.F. McCluer (2004), "Elzinga-Hogarty tests and alternative approaches for market share calculations in hospital mergers", *Antitrust Law Journal* 71:921-947.
- Friedman, J. (1971), "A non-cooperative equilibrium for supergames", *Review of Economic Studies* 38:1-12.
- Froeb, L. and G. Werden (1998), "A robust test for consumer welfare enhancing mergers among sellers of a homogeneous product", *Economic Letters* 58:367-369.
- Fudenberg, D. and E. Maskin (1986), "The folk theorem in repeated games with discounting or with incomplete information", *Econometrica* 54:533-554.
- Fudenberg, D. and J. Tirole (1986), "A 'signal-jamming' theory of predation", *RAND Journal of Economics* 17:366-376.
- Fudenberg, D., D. Levine and E. Maskin (1994), "The folk theorem with imperfect public information", *Econometrica* 62:997-1039.

- Fumagalli, C. and M. Motta (2006), “Exclusive dealing and entry: When buyers compete”, *American Economic Review* 96:785-795.
- Genesove, D. and W. Mullin (2001), “Rules, communication, and collusion: Narrative evidence from the Sugar Institute case”, *American Economic Review* 91:379-398.
- Genesove, D. and W. Mullin (2006), “Predation and its rate of return: The sugar industry, 1887-1914”, *RAND Journal of Economics* 37:47-69.
- Gilbert, R. (1989), “Mobility barriers and the value of incumbency”, in: R. Schmalensee and R. Willig, eds, *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 475-535.
- Gilo, D., Y. Moshe and Y. Spiegel (2006), “Partial cross ownership and tacit collusion”, *RAND Journal of Economics* 37:81-99.
- Green, E. and R. Porter (1984), “Noncooperative collusion under imperfect price information”, *Econometrica* 52:87-100.
- Grossman, S. and O. Hart (1986), “The costs and benefits of ownership: A theory of vertical and lateral integration”, *Journal of Political Economy* 94: 691-719.
- Guzman, A. (1998), “Is international antitrust possible?”, *New York University Law Review* 73:1501-1548.
- Haltiwanger, J. and J. Harrington (1991), “The impact of cyclical demand movements on collusive behavior”, *RAND Journal of Economics* 22:89-106.
- Harrington, J. (1989), “Collusion among asymmetric firms: The case of different discount factors”, *International Journal of Industrial Organization* 7:289-307.
- Harrington, J. (2004a), “Cartel pricing dynamics in the presence of an antitrust authority”, *RAND Journal of Economics* 35:651-673.
- Harrington, J. (2004b) “Post-cartel pricing during litigation”, *Journal of Industrial Economics* 52:517-533.
- Harrington, J. (2005), “Optimal cartel pricing in the presence of an antitrust authority”, *International Economic Review* 46:145-169.
- Harrington, J. (2006a), “Corporate leniency programs and the role of the antitrust authority in detecting collusion”, working paper, John Hopkins University.
- Harrington, J. (2006b), “How do cartels operate?”, available at <http://www.econ.jhu.edu/People/Harrington/fnt06.pdf>.
- Harrington, J. (2007), “Detecting cartels”, in: P. Buccirossi, ed., *Advances in the Economics of Competition Law* (MIT Press, Cambridge), forthcoming.

- Harrington, J. and A. Skrzypacz (2007), "Collusion under monitoring of sales", *RAND Journal of Economics*, forthcoming.
- Hart, O. (1995), *Firms, Contracts, and Financial Structure* (Clarendon Press, Oxford).
- Hart, O. and J. Moore (1990), "Property rights and the nature of the firm", *Journal of Political Economy* 98:1119-1158.
- Hart, O. and J. Tirole (1990), "Vertical integration and market foreclosure", *Brookings Papers on Economic Activity: Microeconomics* 1990:205-286.
- Hastings, J. (2004), "Vertical relations and competition in the retail gasoline market: Empirical evidence from contract changes in Southern California", *American Economic Review* 94:317-328.
- Hausman, J., G. Leonard and C. Velluro (1996), "Market definition under price discrimination", *Antitrust Law Journal* 64:367-386.
- Hausman, J., G. Leonard and J.D. Zona (1994), "Competitive analysis with differentiated products", *Annales d'Economie et de Statistique* 34:159-180.
- Hawk, B. (1988), "The American (anti-trust) revolution: Lessons for the EEC?", *European Competition Law Review* 9:53-87.
- Hayes, J., C. Shapiro and R. Town (2007), "Market definition in crude oil: Estimating the effects of the BP/ARCO merger", *Antitrust Bulletin*, forthcoming.
- Healy, P., K. Palepu and R. Ruback (1992), "Does corporate performance improve after mergers?", *Journal of Financial Economics* 26:255-276.
- Heyer, K. (2006), "Welfare standards and merger analysis: Why not the best?", *Competition Policy International* 2:29-54.
- Holmstrom, B. and J. Tirole (1989), "The theory of the firm," in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 61-133.
- Hovenkamp, H. (1998), *Antitrust Law: An Analysis of Antitrust Principles and Their Application*, vol. 11 (Aspen Law & Business, New York).
- Hylton, K. (2003), *Antitrust Law: Economic Theory and Common Law Evolution* (Cambridge University Press, Cambridge).
- Innes, R. and R.J. Sexton (1994), "Strategic buyers and exclusionary contracts", *American Economic Review* 84:566-584.
- Isaac, R.M. and V. Smith (1985), "In search of predatory pricing", *Journal of Political Economy* 93:320-345.

- Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole (2003a), “The economics of tacit collusion”, Final Report for DG Competition, European Commission, IDEI, Toulouse.
- Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole (2003b), “The economics of unilateral effects”, Interim Report for DG Competition, European Commission, IDEI, Toulouse.
- Jacquemin, A. and M. Slade (1989), “Cartels, collusion, and horizontal merger,” in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 415-473.
- Joskow, P. and A. Klevorick (1979), “A framework for analyzing predatory pricing policy”, *Yale Law Journal* 89:213-270.
- Jung, Y.J., J. Kagel and D. Levin (1994), “On the existence of predatory pricing: An experimental study of reputation and entry deterrence in the chain-store game”, *RAND Journal of Economics* 25:72-93.
- Kandori, M. and H. Matsushima (1998), “Private observation, communication and collusion”, *Econometrica* 66:627-652.
- Kaplan, S. (2000), ed., *Mergers and Productivity* (University of Chicago Press, Chicago).
- Kaplow, L. (1982), “The accuracy of traditional market power analysis and a direct adjustment alternative”, *Harvard Law Review* 95:1817-1848.
- Kaplow, L. (1985), “Extension of monopoly power through leverage”, *Columbia Law Review* 85:515-556.
- Kaplow, L. (2004), “On the (ir)relevance of distribution and labor supply distortion to public goods provision and regulation”, *Journal of Economic Perspectives* 18(4):159-175.
- Kaplow, L. and S. Shavell (1994), “Why the legal system is less efficient than the income tax in redistributing income”, *Journal of Legal Studies* 23:667-681.
- Katz, M. (1989), “Vertical contractual relations”, in: R. Schmalensee and R. Willig eds., *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 655-721.
- Katz, M. and C. Shapiro (2003), “Critical loss: Let’s tell the whole story”, *Antitrust* 17:49-56.
- Kim, E.H. and V. Singal (1993) “Mergers and market power: Evidence from the airline industry”, *American Economic Review* 83:549-569.
- Klemperer, P. (2002), “How (not) to run auctions: The European 3G Telecom auctions”, *European Economic Review* 46(4-5):829-845.
- Klemperer, P. (2004), *Auctions: Theory and Practice* (Princeton University Press, Princeton).
- Kolasky, W. and A. Dick (2003), “The Merger guidelines and the integration of efficiencies into antitrust review of horizontal mergers”, *Antitrust Law Journal* 71:207-251.

- Koller, R. (1971), "The myth of predatory pricing: An empirical study", *Antitrust Law and Economics Review* 3:105-123.
- Kovacic, W., R. Marshall, L. Marx and S. Schulenberg (2006), "Quantitative analysis of coordinated effects", working paper.
- Krattenmaker, T. and S. Salop (1986), "Anticompetitive exclusion: Raising rivals' costs to achieve power over price", *Yale Law Journal* 96:209-293.
- Kreps, D. and J. Scheinkman (1983), "Quantity precommitment and Bertrand competition yield Cournot outcomes", *Bell Journal of Economics* 14:326-337.
- Kreps, D. and R. Wilson (1982), "Reputation and imperfect information", *Journal of Economic Theory* 27:253-279.
- Kühn, K.-U. (2001), "Fighting collusion by regulating communication between firms", *Economic Policy: A European Forum* 16(32):167-197.
- Kwoka, J. and L. White (2004), "Manifest destiny? The Union Pacific and Southern Pacific railroad merger (1996)", in: J.E. Kwoka, Jr. and L.J. White, eds., *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 27-51.
- Lambson, V.E. (1987), "Optimal penal codes in price-setting supergames with capacity constraints", *Review of Economic Studies* 54:385-397.
- Lambson, V.E. (1994), "Some results on optimal penal codes in asymmetric Bertrand supergames", *Journal of Economic Theory* 62:444-468.
- Lambson, V.E. (1995), "Optimal penal codes in nearly symmetric Bertrand supergames with capacity constraints", *Journal of Mathematical Economics* 24:1-22.
- Landes, W. and R. Posner (1981), "Market power in antitrust cases", *Harvard Law Review* 94:937-996.
- Leary, T. (2002), "The essential stability of merger policy in the United States", *Antitrust Law Journal* 70:105-142.
- Levenstein, M. and V. Suslow (2006), "What determines cartel success?", *Journal of Economic Literature* 44:43-95.
- Levin, D. (1990), "Horizontal mergers: The 50 percent benchmark", *American Economic Review* 80:1238-1245.
- Lichtenberg, F. and D. Siegel (1987), "Productivity and changes in ownership of manufacturing plants", *Brookings Papers on Economic Activity* 1987:643-683.

- Litan, R. and C. Shapiro (2002), "Antitrust policy in the Clinton administration", in: J. Frankel and P. Orszag, eds., *American Economic Policy in the 1990s* (MIT Press, Cambridge) 435-485.
- Mankiw, N.G. and M.D. Whinston (1986), "Free entry and social inefficiency", *RAND Journal of Economics* 17:48-58.
- Marshall, R. and M. Meurer (2004), "Bidder collusion and antitrust law: Refining the analysis of price fixing to account for the special features of auction markets", *Antitrust Law Journal* 72:83-118.
- Marvel, H. (1982), "Exclusive dealing", *Journal of Law and Economics* 25:1-25.
- Mason, C., O. Phillips and C. Nowell (1992), "Duopoly behavior in asymmetric markets: An experimental evaluation", *Review of Economics and Statistics* 74:662-670.
- Masten, S. and E. Snyder (1993), "United States versus United Shoe Machinery Corporation: On the merits", *Journal of Law and Economics* 36:33-70.
- McAfee, P. and M. Schwartz, (1994), "Opportunism in multilateral vertical contracting: Nondiscrimination, exclusivity, and uniformity", *American Economic Review* 84:210-230.
- McAfee, P. and M. Williams (1992), "Horizontal mergers and antitrust policy", *Journal of Industrial Economics* 40:181-187.
- McAfee, R.P., H.M. Mialon and M.A. Williams, (2004), "When are sunk costs barriers to entry? Entry barriers in economic and antitrust analysis", *American Economic Review Papers and Proceedings* 94(2):461-465.
- McCutcheon, B. (1997), "Do meetings in smoke-filled rooms facilitate collusion?", *Journal of Political Economy* 105:330-350.
- McGee, J. (1958), "Predatory price cutting: The Standard Oil (N.J.) Case", *Journal of Law and Economics* 1:137-169.
- McGuckin, R. and S. Nguyen (1995), "On productivity and plant ownership change: New evidence from the Longitudinal Research Database", *RAND Journal of Economics* 26:257-276.
- Melamed, A.D. (2006), "Exclusive dealing agreements and other exclusionary conduct: Are there unifying principles?", *Antitrust Law Journal* 73:375-412.
- Milgrom, P. and J. Roberts (1982), "Predation, reputation, and entry deterrence", *Journal of Economic Theory* 27:280-312.
- Morton, F.S. (1997), "Entry and predation: British shipping cartels, 1879-1929", *Journal of Economics & Management Strategy* 6:679-724.

- Motta, M. (2004), *Competition Policy: Theory and Practice* (Cambridge University Press, Cambridge).
- Motta, M. and M. Polo (2003), “Leniency programs and cartel prosecution”, *International Journal of Industrial Organization* 21:347-379.
- Muris, T. (1999), “The government and merger efficiencies: Still hostile after all these years”, *George Mason Law Review* 7:729-752.
- Nalebuff, B. (2003), “Bundling, tying, and portfolio effects, Part 1: Conceptual issues”, Department of Trade and Industry Economics Paper No. 1, available at <http://www.dti.gov.uk/files/file14774.pdf>.
- Nevo, A. (2000a), “Mergers with differentiated products: The case of the ready-to-eat cereal industry”, *RAND Journal of Economics* 31:395-421.
- Nevo, A. (2000b), “A practitioner’s guide to estimation of random-coefficients logit models of demand”, *Journal of Economics and Management Strategy* 9:513-548.
- Nevo, A. (2001), “Measuring market power in the ready-to-eat cereal industry”, *Econometrica* 69:307-342.
- O’Brien, D. and S. Salop (2000), “Competitive effects of partial ownership: Financial interest and corporate control”, *Antitrust Law Journal* 67:559-614.
- O’Brien, D. and G. Shaffer (1992), “Vertical control with bilateral contracts”, *RAND Journal of Economics* 23:299-308.
- O’Brien, D. and A. Wickelgren (2003), “A critical analysis of critical loss analysis”, *Antitrust Law Journal* 71:161-184.
- Office of Fair Trading (2005), “Ex post evaluation of mergers”, March 2005, available at <http://www.of.gov.uk>.
- Ordover, J. and G. Saloner (1989), “Predation, monopolization, and antitrust”, in: R. Schmalensee and R. Willig, eds, *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 537-596.
- Ordover, J. and R. Willig (1981), “An economic definition of predation: Pricing and product innovation”, *Yale Law Journal* 91:8-53.
- Parker, P.M. and L.-H. Röller (1997), “Collusive conduct in duopolies: Multimarket contact and cross-ownership in the mobile telephone industry”, *RAND Journal of Economics* 28:304-322.
- Pautler, P. (2003), “Evidence on mergers and acquisitions”, *Antitrust Bulletin* 48:119-221.

- Perry, M. and R. Porter (1985), "Oligopoly and the incentive for horizontal merger", *American Economic Review* 75:219-227.
- Pesendorfer, M. (2003), "Horizontal mergers in the paper industry", *RAND Journal of Economics* 34:495-515.
- Peters, C. (2003), "Evaluating the performance of merger simulation: Evidence from the U.S. airline industry", *Journal of Law and Economics* 49:627-649.
- Pitofsky, R. (1999), "Efficiencies in defense of mergers: Two years after", *George Mason Law Review* 7:485-493.
- Popofsky, M. (2006), "Defining exclusionary conduct: Section 2, the rule of reason, and the unifying principle underlying antitrust rules", *Antitrust Law Journal* 73:435-482.
- Porter, R. (1983), "A Study of cartel stability: The joint executive committee, 1880-1886", *Bell Journal of Economics* 14:301-314.
- Porter, R. (2005), "Detecting collusion", *Review of Industrial Organization* 26:147-167.
- Porter, R. and J.D. Zona (2004), "Bidding, bid rigging, and school milk prices: *Ohio v. Trauth* (1994)", in: J.E. Kwoka, Jr. and L.J. White, eds. *The Antitrust Revolution: Economics, Competition, and Policy*, 4th ed. (Oxford University Press, New York) 211-232.
- Posner, R.A. (1969), "Oligopoly and the antitrust laws: A suggested approach", *Stanford Law Review* 21:1562-1606.
- Posner, R.A. (2001), *Antitrust Law*, 2nd ed. (University of Chicago Press, Chicago).
- Prager, R. and T. Hannan (1998), "Do substantial horizontal mergers generate significant price effects? Evidence from the banking industry", *Journal of Industrial Economics* 46:433-452.
- Rasmusen, E., M. Ramseyer and J. Wiley (1991), "Naked exclusion", *American Economic Review* 81:1137-1145.
- Ravenscraft, D. and F.M. Scherer (1987), *Mergers, Sell-Offs, and Economic Efficiency* (Brookings Institution, Washington, DC).
- Ravenscraft, D. and F.M. Scherer (1989), "The profitability of mergers", *Journal of Industrial Economics* 7:101-116.
- Rey, P. and J. Tirole (2007), "A primer on foreclosure", in: M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, vol. 3 (Elsevier Science Publishers, New York), forthcoming, available at <http://idei.fr/doc/by/tirole/primer.pdf>.
- Reynolds, R. and B. Snapp (1986), "The competitive effects of partial equity interests and joint ventures", *International Journal of Industrial Organization* 4:141-153.

- Robinson, J. (1933), *The Economics of Imperfect Competition* (Macmillan, London).
- Robinson, M. (1985), "Collusion and the choice of auction", *RAND Journal of Economics* 16:141-145.
- Ross, T.W. (1992), "Cartel stability and product differentiation", *International Journal of Industrial Organization* 10:1-13.
- Rotemberg, J. and G. Saloner (1986), "A supergame-theoretic model of price wars during booms", *American Economic Review* 76:390-407.
- Rothschild, R. (1999), "Cartel stability when costs are heterogeneous", *International Journal of Industrial Organization* 17:717-734.
- Salant, S., S. Switzer and R. Reynolds (1983), "Losses due to merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium", *Quarterly Journal of Economics* 98:185-199.
- Salinger, M. (1990), "The concentration-margins relationship reconsidered", *Brookings Papers on Economic Activity: Microeconomics* 1990:287-321.
- Salop, S. (1986), "Practices that (credibly) facilitate oligopoly co-ordination", in: J. Stiglitz and G. Mathewson, eds. *New Developments in the Analysis of Market Structure* (Macmillan,).
- Salop, S. (2005), "Question: What is the real and proper antitrust welfare standard? Answer: The *true* consumer welfare standard", Statement before the Antitrust Modernization Commission, November 4, 2005.
- Salop, S. (2006), "Exclusionary conduct, effect on consumers, and the flawed profit-sacrifice standard", *Antitrust Law Journal* 73:311-374.
- Salop, S. and D. Scheffman (1983), "Raising rivals' costs", *American Economic Review Papers and Proceedings* 73(2):267-271.
- Scharfstein, D. (1984), "A policy to prevent rational test-market predation", *RAND Journal of Economics* 15:229-243.
- Schelling, T. (1960), *The Strategy of Conflict* (Harvard University Press, Cambridge).
- Scherer, F.M. (1976), "Predatory pricing and the Sherman Act: A comment", *Harvard Law Review* 89:869-890.
- Schmalensee, R. (1989), "Inter-industry studies of structure and performance", in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 2 (North-Holland, Amsterdam) 951-1009.

- Schmalensee, R. (2004), "Sunk costs and antitrust barriers to entry", *American Economic Review Papers and Proceedings* 94(2):471-475.
- Schmalensee, R. and R. Willig, eds. (1989a), *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam).
- Schmalensee, R. and R. Willig, eds. (1989b), *Handbook of Industrial Organization*, vol. 2 (North-Holland, Amsterdam).
- Segal, I. and M. Whinston (2000a), "Exclusive contracts and protection of investments", *RAND Journal of Economics* 31:603-633.
- Segal, I. and M. Whinston, (2000b), "Naked exclusion: Comment", *American Economic Review* 90:296-309.
- Segal, I. and M. Whinston (2003), "Robust predictions for bilateral contracting with externalities", *Econometrica* 71:757-791.
- Selten, R. (1978), "The chain-store paradox", *Theory and Decision* 9:127-159.
- Shapiro, C. (1986), "Exchange of cost information in oligopoly", *Review of Economic Studies* 53:433-446.
- Shapiro, C. (1989), "Theories of oligopoly behavior", in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 329-414.
- Shapiro, C. (1996), "Mergers with differentiated products", *Antitrust*, 10:23-30.
- Simpson, J. and A. Wickelgren (2004), "Naked exclusion, efficient breach, and downstream competition", working paper, University of Texas at Austin, December.
- Smith, A. ([1776] 1970), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Books I-III (Penguin Books, London; first published 1776).
- Snyder, C. (1996), "A dynamic theory of countervailing power", *RAND Journal of Economics* 27:747-769.
- Spence, A.M. (1977), "Entry, capacity, investment and oligopolistic pricing", *Bell Journal of Economics* 8:534-544.
- Spence, A.M. (1981), "The learning curve and competition", *Bell Journal of Economics* 12:49-70.
- Spier, K. and M. Whinston (1995), "On the efficiency of privately stipulated damages for breach of contract: Entry barriers, reliance, and renegotiation", *Rand Journal of Economics* 26:180-202.
- Stigler, G. (1964), "A theory of oligopoly", *Journal of Political Economy* 55:44-61.

- Stole, L. (2007), "Price discrimination and imperfect competition", in: M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, vol. 3 (Elsevier Science Publishers, New York), forthcoming.
- Sutton, J. (2007), "Market structure: Theory and evidence", in: M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, vol. 3 (Elsevier Science Publishers, New York), forthcoming, available at http://personal.lse.ac.uk/SUTTON/market_structure_theory_evidence.pdf.
- Telser, L. (1966), "Cutthroat competition and the long purse", *Journal of Law and Economics* 9:259-277.
- Tirole, J. (1988), *The Theory of Industrial Organization* (MIT Press, Cambridge).
- Tirole, J. (2005), "The analysis of tying cases: A primer", *Competition Policy International* 1:1-25.
- Varian, H. (1989), "Price discrimination", in: R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, vol. 1 (North-Holland, Amsterdam) 597-654.
- Vasconcelos, H. (2005), "Tacit collusion, cost asymmetries, and mergers", *RAND Journal of Economics* 36:39-62.
- Vickers, J. (2005), "Abuse of market power", *Economic Journal* 115: F244-F261.
- Vita, M. and S. Sacher (2001), "The competitive effects of not-for-profit hospital mergers: A case study", *Journal of Industrial Economics* 49:63-84.
- Vives, X. (2001), *Oligopoly Pricing: Old Ideas, New Tools* (MIT Press, Cambridge).
- Waehrer, K. and M. Perry (2003), "The effects of mergers in open-auction markets", *RAND Journal of Economics* 34:287-304.
- Walle de Ghelcke, B. van de and G. van Gerven (2004), *Competition Law of the European Community* (M. Bender, New York).
- Weiman, D. and R. Levin (1994), "Preying for monopoly? The case of Southern Bell Telephone Company, 1894-1912", *Journal of Political Economy* 102:103-126.
- Werden, G. (1996), "A robust test for consumer welfare enhancing mergers among sellers of differentiated products", *Journal of Industrial Economics* 44:409-413.
- Werden, G. (2004), "Economic evidence on the existence of collusion: Reconciling antitrust law with oligopoly theory", *Antitrust Law Journal* 71:719-800.
- Werden, G. (2006), "Identifying exclusionary conduct under Section 2: The 'no economic sense' test", *Antitrust Law Journal* 73:413-433.

- Werden, G. and L. Froeb (1994), "The effects of mergers in differentiated products industries: Logit demand and merger policy", *Journal of Law, Economics and Organization* 10:407-426.
- Werden, G. and L. Froeb (1998), "The entry-inducing effect of horizontal mergers: An exploratory analysis", *Journal of Industrial Economics* 46:525-543.
- Werden, G. and L. Froeb, (2007), "Unilateral competitive effects of horizontal mergers", in: P. Buccirossi, ed., *Advances in the Economics of Competition Law* (MIT Press, Cambridge), forthcoming.
- Werden, G., L. Froeb and D. Scheffman (2004), "A *Daubert* discipline for merger simulation", *Antitrust* 18:89-95.
- Werden, G., A. Joskow and R. Johnson (1991), "The effects of mergers on price and output: Two case studies from the airline industry", *Managerial and Decision Economics* 12:341-352.
- Whinston, M. (1990), "Tying, foreclosure, and exclusion", *American Economic Review* 80:837-859.
- Whinston, M. (2006), *Lectures on Antitrust Economics* (MIT Press, Cambridge).
- White House Task Force Report on Antitrust Policy (1968), in: *Trade Reg. Rep. Supp. No. 415.*, May 26, 1969.
- Williamson, O. (1968), "Economies as an antitrust defense: The welfare trade-offs", *American Economic Review* 58:18-36.
- Williamson, O. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications* (New York, Free Press).
- Williamson, O. (1977), "Predatory pricing: A strategic and welfare analysis", *Yale Law Journal* 87:284-340.
- Williamson, O. (1985), *The Economic Institutions of Capitalism* (New York, Free Press).
- Yamey, B. (1972), "Predatory price cutting: Notes and comments", *Journal of Law and Economics* 15:129-142.
- Zerbe, R. and D. Cooper (1982), "An empirical and theoretical comparison of alternative predation rules", *Texas Law Review* 61:655-715.

Cases

- American Column & Lumber Co. v. United States*, 257 U.S. 377 (1921).
- American Tobacco Co. v. United States*, 328 U.S. 781, 810 (1946).
- Board of Trade of City of Chicago v. United States*, 246 U.S. 231 (1918).
- Broadcast Music, Inc. v. Columbia Broadcasting System, Inc.*, 441 U.S. 1 (1979).
- Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 509 U.S. 209 (1993).
- Brown Shoe Co. v. United States*, 370 U.S. 294 (1962).
- E.I. du Pont de Nemours & Co. v. Federal Trade Commission*, 729 F.2d 128 (2nd Cir. 1984).
- Federal Trade Commission v. Indiana Federation of Dentists*, 476 U.S. 447 (1986).
- Federal Trade Commission v. Staples, Inc.*, 970 F. Supp. 1066 (D.D.C. 1997).
- Federal Trade Commission v. Swedish Match*, 131 F. Supp 2d 151 (D.D.C. 2000).
- Federal Trade Commission v. Tenet Healthcare Corp.*, 17 F. Supp. 2d 937 (E.D. Mo. 1998),
rev'd, 186 F. 3d 1045 (8th Cir. 1999).
- In re High Fructose Corn Syrup Antitrust Litigation*, 295 F.3d 651 (7th Cir. 2002).
- Jefferson Parish Hospital District No. 2 v. Hyde*, 466 U.S. 2 (1984).
- LePage's, Inc. v. 3M*, 324 F.3d 141 (3d Cir. 2003).
- Lorain Journal Co. v. United States*, 342 U.S. 143 (1951).
- Matsushita Electric Industrial Co. Ltd. v. Zenith Radio Corp.*, 475 U.S. 574 (1986).
- Monsanto Co. v. Spray-Rite Service Corp.*, 465 U.S. 752 (1984).
- National Collegiate Athletic Association v. Board of Regents of the University of Oklahoma*, 468 U.S. 85 (1984).
- National Society of Professional Engineers v. United States*, 435 U.S. 679 (1978).
- Spectrum Sports v. McQuillan*, 506 U.S. 447 (1993).
- Standard Fashion Co. v. Magrane-Houston Co.*, 258 U.S. 346 (1922).
- Standard Oil Co. of New Jersey v. United States*, 221 U.S. 1 (1911).
- Standard Oil Co. of California (Standard Stations) v. United States*, 337 U.S. 293 (1949).

Theatre Enterprises v. Paramount Film Distributing Corp., 346 U.S. 537 (1954).

Toys “R” Us, Inc. v. Federal Trade Commission, 221 F.3d 928 (7th Cir. 2000).

United States v. Aluminum Co. of America, 148 F.2d 416 (2d Cir. 1945).

United States v. AMR Corp., 335 F.3d 1109 (10th Cir. 2003).

United States v. Container Corporation of America, 393 U.S. 333 (1969).

United States v. Dentsply International, Inc., 399 F.3d 181 (3d Cir. 2005).

United States v. E.I. du Pont de Nemours & Co., 351 U.S. 377 (1956).

United States v. General Dynamics Corp., 415 U.S. 486 (1974).

United States v. Griffith, 334 U.S. 100 (1948).

United States v. Grinnell Corp., 384 U.S. 563 (1966).

United States v. Microsoft Corp., 56 F.3d 1448 (D.C. Cir. 1995).

United States v. Microsoft Corp., 253 F. 3d. 34 (D.C. Cir. 2001).

United States v. Philadelphia National Bank, 374 U.S. 321 (1963).

United States v. United Shoe Machinery Corp., 110 F. Supp. 295 (D. Mass. 1953),
affirmed per curiam, 347 U.S. 521 (1954).

United States v. Von’s Grocery Co., 384 U.S. 270 (1966).

.***

ICI, Ltd v. Commission, ECR 619, ¶64, Case 48/69 (1972).

Michelin v. European Commission, ECR 3461 § 30, Case 322/81 (1983).

Tetra Pak International SA v. Commission, ECR I-5951, Case C-333/94P (1996).

ECS/Akzo II, Decision of the Commission, 1985 OJ L 374/1 (December 14, 1985).

Hilti, Decision of the Commission, 1988 OJ L 65/19 (December 22, 1987).