Chapter 4

VERTICAL INTEGRATION: DETERMINANTS AND EFFECTS

MARTIN K. PERRY*

Rutgers University
and Bell Communications Research, Inc.

Contents

*This survey was done while the author was a member of the Economic Research Group at Bell Communications Research. This survey represents the views and assumptions of the author and not necessarily those of Bell Communications Research.

Handbook of Industrial Organization, Volume I, Edited by R. Schmalensee and R.D. Willig

"If you want something done right, do it yourself!"
– A theory of vertical integration –
Benjamin Franklin?

1. Introduction

The theory of vertical integration is situated at the intersection of the theory of the firm, the theory of contracts, and the theory of markets. Thus, the literature has developed from several different perspectives. I will discuss the differing perspectives, and comment upon what seem to be the most interesting pieces of the literature. The focus will be on theoretical work, but I will also mention case studies and empirical work when they relate to specific theories. Finally, I will attempt to point out areas of controversy and suggest directions for future research.

1.1. What is vertical integration?

A firm can be described as vertically integrated if it encompasses two single-output production processes in which either (1) the *entire* output of the "upstream" process is employed as *part or all* of the quantity of one intermediate input into the "downstream" process, or (2) the *entire* quantity of one intermediate input into the "downstream" process is obtained from *part or all* of the output of the "upstream" process. This includes the more restrictive criterion that the *entire* output of the upstream subsidiary be employed as *all* of the quantity of one intermediate input into the downstream process. However, both characterizations rule out the case in which *most* of the output of the upstream process is employed as *most* of the input in the downstream process. This case is best described as "partial" vertical integration because some of the output of the upstream process is sold to other buyers and some of the intermediate input for the downstream process is purchased from other suppliers.

Thus, inherent in the notion of vertical integration is the elimination of contractual or market exchanges, and the substitution of internal exchanges within the boundaries of the firm. Suppose that the upstream subsidiary of a firm sold *all* its output to other buyers, and the downstream subsidiary purchased *all* of this intermediate input from other suppliers. We would not describe this firm as being vertically integrated, even if some of the input sold by the upstream subsidiary was eventually resold to the downstream subsidiary. There are no

internal exchanges, only contractual or market exchanges. Instead, this is a vertical "combination".

Although internal exchange provides the observational content of vertical integration, it does not fully capture the essence of vertical integration. Vertical integration also means the ownership and complete control over neighboring stages of production or distribution. In particular, a vertically integrated firm would have complete flexibility to make the investment, employment, production, and distribution decisions of all stages encompassed within the firm. Indeed, the focus of neoclassical models of vertical integration is how these decisions by integrated firms differ from those which would be made by disintegrated firms, or from those which would be socially optimal.

Grossman and Hart (1986) have recently argued that vertical integration is the ownership and thus complete control over "assets". Indeed, they take the position that the nature of the firm's relationship with labor is not relevant for distinguishing vertical integration. The workers could be employees or independent contractors without altering the degree of vertical integration. On the other hand, Williamson (1975) and other authors [see also Cheung (1983)] have emphasized the relationship with labor in discussing vertical integration. For these authors, vertical integration would encompass the switch from purchasing inputs to producing those inputs by hiring labor. The requisite capital, such as buildings and equipment, could be owned or leased without altering the degree of vertical integration. Leasing of capital can allow control of production without ownership. Each view would have certain appeal for specific industries. However, neither view alone provides a complete description of vertical integration. Vertical integration is control over the entire production or distribution process, rather than control over any particular input into that process.

Vertical "controls" characterize vertical relationships between the two extremes of vertical integration and anonymous spot market exchange. A vertical control arises from a contract between two firms at different stages which transfers control of some, but not all, aspects of production or distribution. Examples of vertical controls are resale price maintenance and exclusive territories. Vertical "quasi-integration" is a term coined by Blois (1972) to define financial relationships between firms in neighboring stages. These relationships need not involve additional control of production and distribution decisions. Examples include equity investments, loans or loan guarantees, leases on real estate or capital, and inventory credits.[1] Porter (1980) argues that these arrangements may create a community of interest which can achieve some of the benefits of vertical integration. Vertical control and quasi-integration are intimately

[1]Two specific examples illustrate quasi-integration. IBM has a substantial equity interest in Intel, a leading manufacturer of semiconductors used in IBM products. The major oil refiners often own the service stations that are leased to their franchise dealers.

related to vertical integration. However, since they are the subject of Chapter 11 by Michael L. Katz in this Handbook, we will confine our attention to vertical integration.

Vertical integration may arise in a number of ways. Vertical "formation" describes vertical integration which occurs at the time the firm is created. Vertical "expansion" describes vertical integration which occurs as a result of internal growth of the firm, creating its own new subsidiaries in the neighboring stages. Vertical "merger" describes vertical integration which occurs through the acquisition by one firm of an existing firm in a neighboring stage. In both theoretical and empirical work, the term vertical integration may be used to describe any one of these three patterns. However, the pattern by which vertical integration arises may be intertwined with the determinants and effects of vertical integration.

1.2. Determinants of vertical integration

There are three broad determinants of vertical integration: (1) technological economies, (2) transactional economies, and (3) market imperfections.

Vertical integration may arise from technological economies of integration. In particular, less of the other intermediate inputs may be required in order to obtain the same output in the downstream process when the firm has integrated one of the upstream processes. A typical example is the energy savings from not having to reheat steel in the production of steel sheet. Even though the downstream production process has a well-defined production frontier given a set of intermediate inputs purchased through contracts or markets, a more efficient frontier exists when the set of inputs is broadened to include primary inputs. Vertical integration not only replaces some intermediate inputs with primary inputs, but it also reduces the requirements of other intermediate inputs. This is the sense in which technological economies of integration give rise to vertical integration.

Technological economies may be an important determinant of vertical integration in some industries. However, they will not be a central topic of this chapter. In the theoretical discussions, we will generally presume that firms have integrated so as to internalize technological economies. This allows us to focus upon the more interesting economic reasons for vertical integration. Indeed, Williamson (1975, 1985) has forcefully argued that the costs and hazards of contracting are real reason why "cheek-by-jowl" production processes are generally performed by integrated firms.

Vertical integration may also arise from transactional economies. Transaction costs are different from production costs in that they are associated with the process of exchange itself. However, there is a conceptual analogy to technological economies. One could define the resources consumed in the exchange of

intermediate "technological inputs" as intermediate "exchange inputs" of the downstream production process. Transactional economies could then be defined in exactly the same fashion as technological economies. In particular, vertical integration into the production of intermediate technological inputs would reduce the requirements of intermediate exchange inputs.[2]

Transactional economies are an important determinant of vertical integration. Transactional issues will be prominent in a number of the discussions in this chapter, but only Section 5 will be completely devoted to the "transaction cost economics" of vertical integration. Thus, many of the theoretical discussions will presume that firms have integrated so as to internalize transactional economies. A more general treatment of transaction cost economics is contained in Chapter 3 by Oliver E. Williamson in this Handbook.

Finally, vertical integration can arise from imperfections in markets. Imperfect competition is the most notable example, but other market imperfections also give rise to vertical integration. For example, we will discuss imperfections caused by externalities and imperfect or asymmetric information. How is this distinguished from transaction cost economics?

In transaction cost economics, the primary determinant of vertical integration is "asset specificity" in one or both of the production processes. Asset specificity means that an upstream or downstream firm has made investments such that the value of an exchange is greatest when it occurs between these two firms rather than with other firms. Thus, transaction-specific assets create bilateral monopoly. One cannot then talk about a market for the intermediate input existing between the two firms. Neither firm would necessarily have the ability to set the price or choose the quantity unilaterally. Price, quantity, and any other important dimension of the good (such as quality and delivery) would be determined by negotiation and embodied in a contract. Thus, in transaction cost economics "contractual" exchanges are the relevant alternative to "internal" exchanges. The choice between the two depends upon differences in the cost of "governing" the contractual relationship.

What then is meant by a "market" exchange? These are exchanges which require no negotiations or governance of a continuing relationship. Rather, they are take-it-or-leave-it exchanges in which the price, quantity, and other dimensions of the good are each set by one firm or the other.[3] For example, an upstream manufacturer may set the price and quality, while the downstream

[2] Intermediate exchange inputs could also be integrated, for example by hiring the lawyers necessary for making contracts. However, the economies from such probably arise from the reduced requirement of other exchange inputs, such as expenses for outside legal counsel, rather than technological inputs.

[3] Motty Perry (1986) has recently examined the game theoretic underpinnings of take-it-or-leave-it exchange in a bilateral bargaining situation. With costs of making offers or counter-offers, he finds that the player with the lowest such cost makes the initial offer while the other player accepts or rejects depending upon whether the offer yields a net surplus and does not make a counter-offer.

retailer chooses the quantity to purchase and the delivery schedule. We call this "market" exchange because it is the natural extension of price-taking behavior by firms.

This distinction between contractual exchange and market exchange is artificial and stylized, but it provides a useful method of contrasting transaction cost economics with the neoclassical analysis of vertical integration. The transaction cost analysis of vertical integration discusses the relative advantages of contracts versus internal organization for the joint determination and enforcement of exchange. Various contractual arrangements bridge the gap between vertical integration and anonymous spot markets. The focus is upon the exchange process. On the other hand, the neoclassical analysis of vertical integration assumes that all the relevant dimensions of the good are chosen unilaterally by either upstream firms or downstream firms. The firms make choices so as to maximize their individual profits, rather than the joint profits. Vertical controls then bridge the gap between vertical integration and anonymous spot markets. In this way, the neoclassical analysis of vertical integration or control avoids the bargaining issues of bilateral monopoly by assigning the choice of any particular dimension to the firms at some stage. The focus is upon the production and distribution choices themselves.

Even if firms are integrated to internalize both technological and transactional economies, market imperfections remain an important determinant of vertical integration. Vertical integration in response to technological or transactional economies would generally increase welfare. As a result, transaction cost economics is primarily interested in explaining and predicting patterns of vertical integration. On the other hand, vertical integration in response to market imperfections may increase or decrease welfare. Thus, public policy questions become the primary interest.

1.3. Outline

Sections 2 and 3 discuss neoclassical models in which imperfect competition is the determinant of vertical integration. Section 2 examines forward integration by a monopolist and backward integration by a monopsonist into competitive industries. The incentives for integration are variable proportions, price discrimination, rent extraction, and barriers to entry. Section 3 examines forward integration by a monopolist into a monopolistically competitive industry. The incentives for integration are the elimination of successive markups, internalization of the choice of product diversity, and internalization of service externalities.

Section 4 discusses neoclassical models in which imperfect or asymmetric information is the determinant of vertical integration. The incentives for integra-

tion are diversification, assurance of supply, acquisition of information, and reduction of agency problems.

Section 5 switches to a discussion of vertical integration from the perspective of transaction cost economics. The incentive for vertical integration arises from reducing the transaction costs of bilateral exchange. Evidence for the transaction cost analysis is also summarized and contrasted with the neoclassical interpretations discussed in the previous sections.

Section 6 discusses the theory of incomplete contracts. We first discuss the legal content of an "incomplete" contract, as that term is used by economists. We then examine some recent models which focus upon investments made prior to a bilateral exchange. These models formalize some aspects of the literature on transaction cost economics, and provide a forum for discussing the theoretical distinction between vertical integration and long-term contracts.

Section 7 discusses several related models of "vertical equilibrium". In the same fashion that a horizontal equilibrium defines the output of each firm, a vertical equilibrium would define the stages in which each firm operates. This analysis is at an early stage, so it offers some interesting and challenging questions for research.

Section 8 discusses empirical measures of vertical integration. These measures have been employed to examine aggregate and industry trends in the extent of vertical integration in U.S. manufacturing during the twentieth century. We also discuss some historical cases which shed light upon the pattern of vertical integration prior to the twentieth century.

Section 9 concludes with a discussion of public policy. We summarize the history and current state of antitrust law toward vertical integration in the United States. In the process, we discuss the economic theories which have been the basis for the antitrust challenges to vertical integration. Finally, we briefly discuss public policy toward vertical integration in the United Kingdom and the European Economic Community.

2. Monopoly and monopsony

Imperfect competition at a single stage gives rise to several incentives for the imperfectly competitive firms to integrate into the neighboring competitive stages. These incentives arise from three sources: (1) the internalization of the efficiency losses from the imperfect competitive behavior, (2) the ability to extract inframarginal rents from the competitive stage, and (3) the ability to price discriminate within the competitive stage. We discuss these incentives for both a monopolist integrating forward and a monopsonist integrating backward. However, the insights should clearly apply to oligopolists integrating forward and oligopsonists integrating backward.

2.1. Monopoly and variable proportions

One of the most extensively discussed incentives for vertical integration is that of an upstream monopolist into a competitive downstream industry which employs the monopolist's product in variable proportions with other intermediate inputs. The basic insight is clear. The price set by a monopoly manufacturer leads downstream fabricators to substitute away from the monopoly input toward the other inputs which are competitively supplied. The resulting efficiency loss in downstream production can be converted into profit for the manufacturer when it integrates into the fabrication stage and expands the usage of its input to the efficient level. Note that no such incentive exists when the manufacturer's product is used in fixed proportions with other inputs by a constant cost downstream stage. In this case, the final demand is simply mirrored to the manufacturer by the competitive downstream stage, and the manufacturer can maximize industry profits without forward integration.[4]

The incentive created by variable proportions was first discussed by McKenzie (1951). However, a brief graphical illustration by Vernon and Graham (1971) triggered a steady stream of research on this model since then. The driving force in this literature has been the welfare question of whether vertical integration by the monopoly manufacturer will increase or decrease welfare. The loss from inefficient production is eliminated and captured by the manufacturer in the form of profits. The size of this gain depends upon the elasticity of substitution in production. However, the integrated manufacturer may increase the retail price of the fabricated product and thereby reduce consumer welfare. This loss depends in part upon the elasticity of demand. Thus, there are two issues. First, when is the retail price higher after integration? And second, if the retail price is higher, is it so high that total welfare (consumer plus producer surplus) actually decreases?

Schmalensee (1973) showed that the retail price rises when the retail production function is Cobb–Douglas and demand has a constant elasticity (greater than unity). But even in this case, he was not able to draw any general conclusions about the direction of total welfare. Using a constant elasticity of substitution (CES) production function, Hay (1973) showed that the retail price rises for many values of the elasticity of substitution. In particular, the retail price rises if the elasticity of substitution exceeds the elasticity of demand. Computations by Warren-Boulton (1974) indicated that even if this condition is not satisfied, the retail price may still rise. Warren-Boulton also found that welfare could increase even though the retail price rises. But the welfare increases

[4]Quirmbach (1986a) demonstrates that there can be "scale" incentives for the manufacturer to forward integrate if the downstream stage is composed of firms with U-shaped average cost curves. The monopoly price of the input alters the minimum efficient scale of downstream firms and causes a distortion in the number of firms. Forward integration by the manufacturer can correct this.

are relatively small and occur only for values of the elasticity of substitution less than some value itself less than unity. This would appear to be counterintuitive in that the efficiency gains increase with the elasticity of substitution (up to a point). But the increase in the retail price also becomes more significant when the elasticity of substitution is larger. The substitution possibilities at the retail stage are an important constraint on the wholesale pricing of the manufacturer.

Other authors have generalized this model in some fashion. Mallela and Nahata (1980) re-examine the CES production function and derive the condition necessary for the final price to decline when the elasticity of substitution is less than unity. Westfield (1981) employs a more general framework to summarize all of the retail price results. He finds specific cases in which the retail price falls when the elasticity of substitution is less than unity. Waterson (1982) employs the CES production function, but assumes that the downstream industry is a Cournot oligopoly rather than competitive. Finally, Quirmbach (1986b) assumes a competitive fringe at the upstream stage and examines forward integration by the now dominant firm.

This literature generates a few policy conclusions. First, Westfield's summary suggests that we should not expect the retail price to fall when manufacturers integrate into subsequent fabrication stages of the industry. Second, the incentive to integrate is relatively strong in that the manufacturer increases its profits both by internalizing the efficiency gain and by setting a higher retail price. But third, the reduction in total welfare will be relatively small. Although Warren-Boulton did not calculate the percentage welfare loss, his results seem to suggest that it is less than a couple of percent. Moreover, oligopoly would generate smaller losses than monopoly. Thus, it is not clear that variable proportions raises a major policy issue on vertical integration.

Vertical integration in response to variable proportions is a relatively drastic solution to what is primarily a pricing problem. Blair and Kaserman (1983) have shown that tying arrangements, output or sales royalties, and lump-sum entry fees can all be used by the monopolist to eliminate and internalize the efficiency loss from simple monopoly pricing. Warren-Boulton (1977) examines the use of these alternatives by the United Mine Workers Union, the monopolist over the supply of labor to a competitive mining industry. Thus, the variable proportions model is probably a better description of some observed vertical controls than of vertical integration.

2.2. Monopoly and price discrimination

The separation of downstream markets for price discrimination is perhaps a more important incentive for forward integration by manufacturers. Wallace (1937)

seems to have been aware of this point in his pioneering case study of Alcoa, but Stigler (1951) clearly stated the basic argument [see also Gould (1977)].

Suppose a monopoly manufacturer is selling its product to two competitive fabrication stages, one with an elastic derived demand and one with an inelastic derived demand. If the manufacturer fully integrated both downstream stages, he would contract input employment in the inelastic stage and expand input employment in the elastic stage. But the manufacturer could also duplicate this outcome by integrating into only the elastic stage. He would expand input employment in that stage while raising the input price to the monopoly price of the inelastic stage. This strategy results in a "price squeeze" against independent fabricators in that the manufacturer's downstream subsidiary lowers the price of fabrications while raising the market price of the input.[5] Because of this, the monopolist can effectuate price discrimination by vertical expansion rather than vertical merger.

If the manufacturer reversed this strategy by integrating into the inelastic stage, resale and free entry would prevent the manufacturer from achieving price discrimination. The manufacturer could not contract input employment in the inelastic stage while lowering the input price intended for the elastic stage. The higher retail price in the inelastic stage would be undercut by new entrants who obtain the input at the lower market price either from the manufacturer (if he cannot identify buyers) or from the fabricators in the elastic stage on resale.

Although price discrimination is an important determinant of vertical integration, no clear policy conclusions can be drawn from this analysis. Third-degree price discrimination may increase or decrease total welfare depending upon the nature of the derived demands of the downstream industries.[6]

The price discrimination story becomes more interesting if the manufacturer faces some competition at his stage of production. Perry (1978a) discusses the simplest such case. There are many competitive fabrication stages, and they can obtain the necessary input from either the dominant manufacturer or from a competitive fringe of small manufacturers. Perry then examines the extent to which the dominant manufacturer can still price discriminate in the presence of the fringe. By the previous reasoning, the optimal strategy for dominant manufacturer is to integrate into a set of stages which have more elastic derived demands than the remaining non-integrated stages. Since these are the stages with the lower monopoly prices, the stages can be ordered by their monopoly prices. Let $\hat{R}_j$ be the monopoly price of the jth fabricating stage, such that $\hat{R}_1 < \hat{R}_2 < \cdots < \hat{R}_m$. When the dominant manufacturer integrates into one of the stages, that demand is eliminated from the input market in which the fringe

[5]Carlton and Perloff (1981) examine the implications of this theory for depletion of a natural resource when the monopolist is the sole owner of the resource.

[6]See Schmalensee (1981) and Varian (1985).

manufacturers compete. As a result, the dominant manufacturer optimally resets the open market price of its product, taking into account (1) the lower but less elastic demand and (2) the unchanged supply curve of the fringe manufacturers. Let $\hat{P}_j$ be profit-maximizing open market price when all industries $i > j$ are not integrated.

If the fringe is not too large, $\hat{P}_0 > \hat{R}_1$ since $\hat{P}_0$ is the dominant firm's price to all the downstream industries while $\hat{R}_1$ is the monopoly price to the industry with the least elastic demand. For small j, $\hat{P}_j$ increases because the market demand is less elastic. But for large j, $\hat{P}_j$ decreases because the market demand is small relative to the fringe supply. Perry then defines the industry $\xi = \max\{ j\colon \hat{R}_j < \hat{P}_j \}$. The dominant firm can integrate the downstream industries up to and including the ξth industry without creating resale possibilities at the profit-maximizing prices. However, forward integration beyond the ξth industry requires the dominant firm to contract open market sales in order to keep the open market price above the profit-maximizing internal prices of the additional integrated industries. Otherwise, new entrants could undercut the final price in these industries. Perry shows that the gains from setting the monopoly price in an industry $j > \xi$ are more than offset by the reduction in profits from open market sales. Thus, ξ is the maximum number of downstream industries over which the dominant firm can price discriminate. Moreover, ξ decreases with outward shifts in the supply curve of the competitive fringe.[7]

Perry (1980) applies this theoretical analysis to the forward integration by Alcoa prior to 1930. Alcoa was the only domestic producer of primary aluminum during this period, but there existed several foreign producers whose exports to the United States were subject to a tariff which varied over the period. Thus, the industry arguably approximated the dominant firm model, raising the question whether the observed pattern of forward integration into domestic fabrication stages by Alcoa was consistent with the motive of price discrimination. Consider four of the major uses for aluminum during this period: (1) cooking utensils, (2) electric cable, (3) automobile parts, and (4) aircraft parts. Using rough judgments, Perry estimates that the derived demand for aluminum by each of these industries. The derived demand was very elastic for the electric cable industry, and very inelastic aircraft industry. However, the elasticity varied for different products in the cooking utensil and automobile part industries depending upon the other materials which were good substitutes (primarily iron and steel, but also secondary aluminum).

[7]Oligopolists could also price discriminate by forward integration in the same fashion. But it is not clear whether dominant oligopolists would be able to integrate more or less of the elastic industries than a dominant firm. The oligopoly input price is lower, but the oligopoly final prices in the integrated industries would also be lower. The oligopolists could obviously do even better if they could each fully integrate and individually monopolize a subset of the elastic industries.

The pattern of forward integration by Alcoa appears to be roughly consistent with price discrimination. There was complete integration into the fabrication of aluminum cable, very little integration into the production of aircraft parts, and selective integration into those cooking utensils and automobile parts which were most subject to competition from other materials.[8] These conclusions are crude, but they raise questions about the traditional explanation that Alcoa integrated forward only after it was unable to persuade independent metal fabricators to use aluminum in the production of their goods.[9]

McNicol (1975) has a related story of vertical integration and price discrimination. McNicol argues that a manufacturer who is partially integrated into one of the downstream industries with an inelastic demand may increase its profits by rationing the intermediate input to the independent firms in that downstream industry. The manufacturer cannot explicitly raise the price of the intermediate input to such industries, but he can raise the final price of such industries by rationing the independent producers. Partial integration into these industries then allows the manufacturer to share in their higher profits. But the increase in profits of the downstream subsidiary must dominate the reduced profits on sales to the independent firms at the unchanged market price of the intermediate input. McNicol calls this "quantity discrimination". Unlike price discrimination, quantity discrimination implies that producers would partially integrate into those industries with inelastic demands, rather than the industries with elastic demands.[10] But like price discrimination, quantity discrimination results in a "supply squeeze" of independent fabricators.

Vertical integration for quantity discrimination has problems which are not present in vertical integration for price discrimination. If resale is costless, it would limit the ability of a producer to effectively ration independent fabricators in an industry with an inelastic demand. Independent fabricators could obtain the input on the spot market. But suppose resale is costly. If the manufacturer can identify buyers, as he must in order to ration any particular industry, then why cannot the producer price discriminate with a different input price to each downstream industry? The Robinson–Patman Act may be a barrier to such explicit price discrimination. Alternatively, McNicol may be implicitly assuming

[8]Alcoa's forward integration into aluminum sheet, cookware, and automobile parts was investigated by the FTC during most of the 1920s. Independent firms in those industries frequently alleged that they were unprofitable and went out of business because of Alcoa's integration and subsequent price squeeze. See Wallace (1937).

[9]See Peck (1961) for a discussion of the traditional explanation forward integration by aluminum producers. Silver (1984, ch. 12) disputes Perry's conclusions and argues in favor of the traditional explanation.

[10]Curiously, McNicol argues that the copper wire industry had important substitution possibilities, suggesting that its demand was elastic. However, this fabricating industry was partially integrated and rationed by the copper producers during the period 1946–70.

that the producer is only one member of an upstream oligopoly and that oligopoly pricing inhibits price discrimination.

These models posit forward integration as an implicit method of achieving price discrimination. However, Katz (1987) points out that the threat of backward integration may be the cause of explicit price discrimination in intermediate good markets. Katz examines a model in which a retail chain competes with local firms in geographically separated markets. The chain and the local retailers purchase inputs from an upstream monopolist. Because of scale economies in upstream production, the chain has a stronger incentive to integrate backward. If the upstream monopolist cannot price discriminate, this threat results in lower input prices for both the chain and its local competitors. However, if the upstream monopolist can price discriminate, it can increase its profit by charging a higher input price to the local firms. This will benefit the chain by giving it a cost advantage over its local competitors. As a result, the upstream monopolist can even raise the input price to the chain slightly without inducing it to integrate backward. The threat of backward integration makes the derived demand of the chain infinitely elastic at some price. This suggests that implicit price discrimination could also be effectuated by a vertical merger between the upstream monopolist and the retail chain.

2.3. Monopsony and backward integration

Incentives also exist for backward integration by a monopolist into an upstream competitive stage. McGee and Bassett (1976) discuss the basic efficiency incentive. Consider a manufacturer who is the sole buyer of a raw material which is supplied by competitive firms but subject to a rising supply price. Because of the rising supply price, the expenditure of the manufacturer for an additional unit of the raw material exceeds the supply price. Thus, the manufacturer employs too little of the raw material. Vertical integration eliminates this monopsony inefficiency. Industry profits and welfare obviously increase.

Perry (1978c) expands this model to include the process of vertical merger by the monopsonist. Backward integration by the manufacturer requires that the assets of the raw material suppliers be acquired at some price exceeding the rents they were earning. The manufacturer could initially acquire all of the suppliers, paying them the rents earned under monopsony. However, the manufacturer can actually extract some of these initial rents by acquiring the suppliers one at a time. Cost minimization by a partially integrated manufacturer results in an expansion in the production of the subsidiaries and a reduction in the purchases from the remaining independent suppliers. This reduces the market price of the raw material and the rents to independent suppliers. Thus, the manufacturer can acquire new subsidiaries at a lower price. Moreover, such an acquisition strategy

is credible because the manufacturer is maximizing profits at each point. With an optimal acquisition program, the manufacturer can substantially reduce the cost of integration. This extraction of rents is a form of imperfect price discrimination which enhances the manufacturer's incentive to integrate backward.

Backward integration by a dominant manufacturer may also create a barrier to entry so as to preserve its dominance. Bain (1956) popularized the concept of barriers to entry and also discussed the importance of potential competition. Bain argued that vertical integration creates a capital barrier to entry by forcing potential entrants to contemplate entry at two stages of production rather than just one. In addition, he pointed out that vertical merger also eliminates one of the most natural potential entrants into each stage. Indeed, these two theories are complements. It is difficult to argue that firms in neighboring stages are the most likely entrants without also believing that entry at both stages is more difficult than entry at one stage.

Similarly, a dominant firm may also use vertical integration to raise the costs of its competitors. Salop and Scheffman (1983) have discussed vertical integration as one device which can be used to "raise rivals' costs".[11] An extreme example is the acquisition of essential inputs or facilities which would genuinely foreclose competitors.[12] The opportunity cost of holding such resources may be less than the risk of reduced profits from new competition. A dominant firm need not acquire all of the scarce resource in order to effectively raise the costs of its actual or potential rivals. By leaving the open market thin, competitors may be unable to expand without significantly driving up the input price, they may be subject to higher prices set by the fewer remaining suppliers, or they may incur higher transaction costs from having to negotiate contracts with suppliers.

In an attempt to illustrate the second of these three possibilities, Ordover, Saloner and Salop (1987) examine a model of successive duopoly with price-setting. By acquiring one of the upstream firms, a downstream firm can disadvantage and reduce the profits of its competitor by forcing it to rely on an upstream monopolist for the input. Moreover, the integrated firm can prevent its competitor from merging with the upstream monopolist by continuing to offer the input for sale at some appropriate price less than the monopoly price. However, if both downstream firms can bid to acquire one of the upstream firms, a prisoner's

[11]Such devices would also include vertical controls such as exclusive dealing, or horizontal restraints such as exclusionary contracts or product standards. They would also include political and regulatory activity. Such devices can be less costly than acquiring competitors or engaging in predatory pricing to drive them out of business.

[12]The most notable historical example was the formation of the Terminal Railroad Association of St Louis to monopolize the ferry and bridges across the Mississippi River connecting eastern and western railroad lines. The facilities were for exclusive use of the member railroads. The Supreme Court held the combination to be a violation of the Sherman Act, and required the Association to admit members and set prices in a non-discriminatory fashion. *U.S. v. Terminal Railroad Association* (1912, 224 U.S. 383).

dilemma arises in that the winning bid makes the integrated firm less profitable than it was when both downstream firms were disintegrated.

It is useful to consider the early domestic steel and aluminum industries in light of these alternative theories of backward integration.

Mancke (1972) examined the rapid backward integration into iron ore production by Carnegie and the other large steel producers during the period 1896–1900 before the formation of U.S. Steel. In 1896, two-thirds of the iron ore consumed in the United States came from the Lake Superior region. Of the more than one hundred iron ore suppliers in this region, only one produced as much as 10 percent of the ore. Moreover, only Carnegie Steel had an interest in any of these suppliers, and that interest was insignificant. But by 1900, eight steel producers completely controlled almost three-quarters of this region's output, each becoming self-sufficient in iron ore. Moreover, this backward integration coincided with declining prices for both iron ore and iron ore properties. Mancke concludes that the steel producers were less pessimistic about the trend of future iron ore prices than the owners of iron ore properties. Iron ore prices did rise abruptly thereafter. In 1901, U.S. Steel was formed by a merger of Carnegie Steel and other major steel producers. As its predecessors, U.S. Steel continued to acquire iron ore properties. Parsons and Ray (1975) document these events and argue that control of iron ore deposits by U.S. Steel inhibited domestic entry and preserved the dominance of U.S. Steel longer than would have been the case otherwise. Given the rapid increase in concentration during this period, subsequent backward integration by the steel producers is also not inconsistent with the monopsony theory.

During its early years as the domestic monopolist of primary aluminum, Alcoa was completely integrated into the extraction of bauxite and the generation of electricity for refining alumina into aluminum. By 1910, Alcoa had acquired most of the bauxite lands in Arkansas, shipping the ore to its sole alumina plant in East St. Louis. When the Arkansas ore began to deplete, Alcoa turned to South America acquiring or leasing nearly all the deposits in the Guianas by the end of World War I. Alcoa built and operated hydroelectric power stations near its refineries in the United States. These acquisitions are perhaps a classic example of Alcoa erecting entry barriers in light of the expiration of the domestic patents on the aluminum reduction process. However, they are also consistent with the monopsony theory.

Crandall (1968) proposes an interesting model to explain backward integration into the production of automobile parts by the automobile manufacturers. The argument could apply to the assembly stage of any industry in which there would be demand for replacement parts on the durable final good. If the assembler could not integrate backward, a competitive parts industry increases the demand for the final good by lowering the price of replacement parts for consumers. However, if the assembler can integrate backward and monopolize the parts industry, then he would clearly charge a higher price to consumers for replace-

ment parts even though it caused some reduction in demand for the durable good. This would price discriminate against consumers who use the durable good more intensively. This is the common explanation for tying arrangements [see Burstein (1960)]. Thus, backward integration allows price discrimination in the sense of optimal pricing of complementary goods [see Telser (1979)].

The assembler can certainly integrate the production of parts for assembly of the durable good. But because of the high price of replacement parts, other vertical controls would generally be necessary to prevent entry into parts. Crandall discusses some of the practices used by the automobile assemblers to inhibit such entry. Crandall then finds that the automobile assemblers are highly integrated into a number of components. Moreover, the assemblers seem to make very high returns on the replacement parts. This strategy would not be available for replacement parts, such as tires, which consumers clearly recognize as not being specialized to their automobile.

3. Monopolistic competition

In this section we focus upon the incentives created by successive stages of imperfect competition. Price remains an important dimension of the analysis. However, we focus upon the problems raised by new dimensions such as product differentiation and retail service externalities. Imperfect competition with the disintegrated structure prevents the industry from achieving the profit-maximizing price, diversity, or service. Vertical integration corrects these choices for the industry but it need not increase total welfare.

Consider the simplest case of successive monopolists, first examined by Spengler (1950) and Machlup and Taber (1960). A manufacturer sells to regional wholesalers who sell to local retailers. The retailers have local monopolies, the wholesalers have regional monopolies, and the manufacturer has a national monopoly. With constant marginal costs of distribution, the marginal revenue function of a retailer becomes that retailer's inverse demand function to the wholesaler. Similarly, the marginal revenue function of a wholesaler's aggregate inverse demand function is the manufacturer's inverse demand function from that wholesaler. Thus, each monopoly stage rotates the inverse demand function downward and causes the upstream monopolist, here the manufacturer, to produce an output less than the output that would maximize industry profits. Vertical integration by the manufacturer into all stages of distribution would reduce the final price, thereby increasing industry profits and consumer welfare.[13]

[13]The analysis and results would obviously be identical for forward integration by upstream oligopolists into a downstream oligopoly. See Greenhut and Ohta (1976), the comments by Perry (1978b) and Haring and Kaserman (1978), reply by Greenhut and Ohta (1978), and subsequent correction by Greenhut and Ohta (1979).

3.1. *Monopolistic competition at the retail stage*

The price effects of successive monopoly provide a clear prescription for vertical integration when the product is homogeneous and the wholesalers or retailers have isolated monopolies. However, successive monopoly raises a more interesting and difficult issue when the retail stage is monopolistically competitive. When retailers are differentiated and compete for consumers at the margin, we are concerned not only with the final price but also with the number of retailers. The ease of access to a retailer or the diversity in the choice among retailers affect consumer welfare. Thus, we must examine the impact on the number of retailers before we can prescribe vertical integration as a remedy for successive monopoly. This issue has been discussed in recent work by Dixit (1983), Mathewson and Winter (1983), and Perry and Groff (1985).

The models posit a manufacturer whose product is distributed to consumers by a retail stage. The manufacturer is a simple monopolist, but the retail stage is monopolistically competitive. The product of each retailer is differentiated, but free entry increases the number of retailers to the point where each earns no profit. Dixit (1983) and Mathewson and Winter (1983) employ the spatial model of retail differentiation, popularized by Salop (1979). On the other hand, Perry and Groff (1985) employ the CES model of retail differentiation, introduced by Spence (1976) and Dixit and Stiglitz (1977).

In the spatial model of Salop (1979), L consumers are uniformly distributed along a circular market of unit length, and must incur a travel cost t per unit distance in order to reach a retailer. Each retailer has a fixed cost f, but its only variable cost is the wholesale price r of the manufacturer's good. The manufacturer is a monopolist and has a constant cost c of producing the good. Salop assumes that consumers purchase only one unit of the good, and that they have a reservation price v. Consumers then purchase the good from the retailer with the lowest full price (retail price plus transportation costs) as long as that full price is less than their reservation price. Thus, the price set by each of two neighboring retailers will determine the market boundary where consumers are just indifferent between the two retailers. As a result, each retailer faces a locally linear demand function, given the price set by the two retailers on either side. In particular, a lower price reduces the retailer's full price to all consumers, causing consumers at the boundaries to shift their purchases to him. This is the sense in which each retailer is a monopolist, differentiated from its neighboring retailers, yet in direct price competition with them.

Given a low wholesale price, a Nash equilibrium can be defined in which each retailer sets its price assuming that the prices of neighboring retailers will remain unchanged. Free entry then determines the total number of retailers n. This is Salop's "competitive equilibrium", and all consumers clearly purchase the good from some retailer. At some higher wholesale price this "competitive equilibrium"

will not exist. Instead, free entry drives the industry to the highest price and lowest output such that the profits of all retailers are just non-negative. This is Salop's "kinked equilibrium". Finally, at a sufficiently high wholesale price, $r_m = v - [2 \cdot t \cdot f/L]^{1/2}$, there remains only one such price–output combination which enables retailers to earn non-negative profits. This is Salop's "monopoly equilibrium", so called because it occurs at the output where the monopoly price of a retailer equals its average cost.

At wholesale prices above r_m, retailers cannot earn non-negative profits at any price, and sales of the good fall to zero. Ignoring the integer problem on the number of retailers, free entry insures that all consumers purchase the good in each equilibria with $r < r_m$. Thus, the derived demand facing the manufacturer is perfectly inelastic up to the wholesale price r_m, which then becomes the profit-maximizing wholesale price. The monopoly equilibrium prevails and the number of retailers is $n_m = [tL/2f]^{1/2}$. Since all consumers purchase the good, there is no efficiency loss caused by successive monopoly pricing, and thus no corresponding incentive for the manufacturer to vertically integrate. The free entry number of retailers is also the number that would be chosen by an integrated manufacturer.[14] Thus, there is no distortion in the retail diversity from the monopoly equilibrium, and therefore no diversity incentive for vertical integration.

Salop's model generates no price or diversity incentive for forward integration by the manufacturer. The reason is that the derived demand facing the manufacturer reflects the perfectly inelastic demand of each consumer up to the reservation price. Dixit (1983) avoids this degeneracy by assuming that retailers actually produce the final good using the manufacturer's product in variable proportions with other inputs. If the elasticity of substitution is sufficiently high, the profit-maximizing wholesale price will be less than r_m. This would then leave the retail stage in a kinked or competitive equilibrium with a higher final price and a larger number of retailers than the monopoly equilibrium. The manufacturer now has an incentive to integrate, and welfare will increase with vertical integration. Vertical integration eliminates the inefficient production at the retail stage. Moreover, the integrated manufacturer reduces the number of retailers to n_m, which is closer to the optimal number of retailers n^*. Moreover, since consumption of the final good remains unchanged, any change in the retail price is merely a transfer with no deadweight loss.

Mathewson and Winter (1983) maintain fixed proportions in retailing, and instead generalize consumer demand. They assume that each consumer has an

[14] To see this, consider the profit function of the integrated manufacturer. Since the manufacturer would either increase the number of retailers or decrease the retail price in order to sell his good to all consumers, we can let $p = v - t/2n$, and write the manufacturer's profit function solely in terms of n as $\Pi(n) = (v - t/2n - c) \cdot L - n \cdot f$. The number of retailers that maximize this profit function is n_m, the same number of retailers as in the monopoly equilibrium with r_m as the wholesale price.

exponential demand for the good. But in so doing, they must also assume that the transportation cost t is per unit of the product purchased as well as per unit distance traveled. An alternative interpretation is that the consumers at each location still demand only one unit of the good, but that they have different reservation prices which would aggregate to an exponential demand. Let p_j be the price of the jth retailer and p be the price of the neighboring retailers. The demand function facing the jth retailer can be expressed as

$$x_j(p_j, p, n) = (2v/t) \cdot [1 - e^{(p_j - p - t/n)/2}] \cdot e^{-p_j}. \quad (1)$$

The derived demand facing the manufacturer at any given wholesale price is obtained from the free entry Nash equilibrium in prices at the retail stage. The profit-maximizing wholesale price will result in both a pricing distortion and a diversity distortion.

Vertical integration allows marginal cost internal pricing to the retail subsidiaries. This results in a lower final price, eliminating the pricing distortion and increasing welfare. However, vertical integration also reduces the number of retailers, by allowing the manufacturer to internalize the entry externality. Without integration, a potential entrant into retailing would ignore the reduction in the profits of other retailers when making its decision to enter. But the integrated manufacturer will take this reduction into account when considering a new retailer. As a result, the integrated manufacturer will always reduce the number of retailers from that which would arise in the free entry disintegrated equilibrium. The resulting welfare evaluation of vertical integration will then depend upon whether there is a welfare loss from the reduction in diversity, and if so, how large that welfare loss is.

In the spatial model used by Dixit, the free entry equilibrium yields excessive retail diversity, and vertical integration produces a welfare gain by reducing the number of retailers. However, this is not the case in the spatial model used by Mathewson and Winter. Their simulations indicate that the optimum ($p = c$) and the constrained optimum ($r = c$) both call for *more* retailers than the free entry Nash equilibrium under wholesale pricing. Thus, vertical integration generates a welfare loss from reduced retail diversity. However, this welfare loss is dominated by the welfare gain from eliminating successive markups. Mathewson and Winter discover small welfare gains from vertical integration.

The welfare conclusion of Mathewson and Winter can be reversed when a non-spatial model of retail differentiation is used. Perry and Groff (1985) examine a convenient non-spatial model. In particular, they employ the constant elasticity of substitution (CES) benefit function introduced by Spence (1976) to generate the derived demands for retailers. If x_i is the output of the ith retailer, a composite commodity y can be defined as $y = \sum \alpha \cdot x_i^{\beta}$, where $\alpha > 0$ and $0 < \beta < 1$. The benefit function is then specified as $U(\cdot) = y^{\theta}$, where $0 < \theta < 1$. Since the elasticity of substitution is $\sigma = 1/(1 - \beta)$, $\beta < 1$ insures that the retailers are differentiated. Assuming no income effects, the inverse demand

function facing the kth retailer can be expressed as

$$p_j(x_1, \ldots, x_n) = \alpha\beta\theta \cdot \left\{ \sum_{i=1}^{n} \alpha \cdot x_i^{\beta} \right\}^{\theta-1} \cdot x_j^{\beta-1}. \tag{2}$$

With this demand structure, Perry and Groff define a free entry "monopolistically competitive" equilibrium in which each retailer chooses its quantity assuming that the weighted average of all outputs remains unchanged. This equilibrium generates a derived demand for the manufacturer's good. The profit-maximizing wholesale price of the manufacturer generates both a price and a diversity distortion. Vertical integration by the manufacturer eliminates the wholesale markup and lowers the retail price. Vertical integration also reduces the number of retailers, moving the retail stage farther from the optimum number of retailers. But unlike the Mathewson and Winter model, the welfare loss to consumers from reduced retail diversity now dominates the gains from eliminating the wholesale markup for all values of the parameters. Thus, vertical integration reduces welfare. This also occurs when there are oligopolists at the manufacturing stage. Oligopoly only lessens the severity of the welfare loss.

These models make it clear that the welfare assessment of vertical integration by a manufacturer depends upon the nature of consumer demand facing the monopolistically competitive retail stage. Consider two differences between the spatial and CES models. First, competition is "localized" in the spatial model, but "generalized" in the CES model. In (1), the quantity demanded from the jth retailer depends only upon the price p charged by the neighboring retailers. On the other hand, the demand price for the jth retailer in (2) depends upon the quantity sold by all other retailers. Thus, each retailer competes equally with all other retailers in the CES demand model. This suggests that we might expect more entry in the spatial model, implying that the free entry equilibrium may not perform too badly with respect to retail diversity. Second, new retailers "crowd" the market in a spatial model, whereas they create a new dimension to the product space in the CES model. In other words, entry does not increase total demand in the spatial model, but it does in a CES model. This suggests that the spatial model would place a lower value on retail diversity than the CES model. These two considerations may account for the fact that reduced retail diversity does not prevent vertical integration from increasing welfare in the model of Mathewson and Winter, but it does in the model of Perry and Groff.

3.2. Monopolistic competition and externalities at the retail stage

The welfare assessment of vertical integration by a manufacturer into retailing will also depend upon yet other non-price characteristics of the retail stage that

enter into consumer benefits. In particular, Mathewson and Winter (1984, 1986) and Perry and Porter (1986) provide further extensions of the spatial and CES models to address advertising or service by retailers. The primary purpose of these papers is to examine vertical controls, but they also have important implications for vertical integration.

Mathewson and Winter (1984) use the same spatial model as their previous paper, but introduce the additional complication that retailers must advertise in order to inform consumers about the existence of the product. An increase in the advertising of one retailer will inform more consumers in that retailer's market area, but it will also inform consumers in the market areas of other retailers. This advertising spillover dulls the incentive of retailers to advertise, primarily to the detriment of the manufacturer. Thus, vertical integration will increase advertising. Mathewson and Winter (1986) are not able to make definitive conclusions about the retail price and the number of retailers, but some calculations indicate that vertical integration generally reduces the number of retailers, but does not necessarily reduce the retail price. Moreover, their calculations indicate that welfare increases with vertical integration, as in their model without the externality. Since advertising is purely informative and brings new consumers into the market, it would seem that the increased retail advertising would benefit both the manufacturer and consumers.

Similarly, Perry and Porter (1986) generalize the CES model of Perry and Groff (1985) to allow for service externalities in retailing. The composite commodity y is modified so that it depends upon both the quantity purchased, x_i, and the service received, z_i. In particular, $y = \sum z_i^{\alpha} \cdot x_i^{\beta}$, where $\alpha + \beta < 1$. Service can be thought of as presale non-appropriable information. However, the service received by a consumer is assumed to be a linear combination of the service of the retailer from whom the good is purchased and the average level of service by all retailers, i.e. $z_i = \lambda \cdot s_i + (1 - \lambda) \cdot \bar{s}$, where $0 \leq \lambda \leq 1$. Thus, there is an externality which reduces the incentive of retailers to perform service. This reduces the derived demand for the manufacturer's good and limits the profits that can be achieved with wholesale pricing. Vertical integration allows the manufacturer to internalize the externality, increase the retail service, and increase profits. However, welfare need not increase. Since vertical integration reduces welfare in the model of Perry and Groff without a service externality, it is not surprising that Perry and Porter find that vertical integration still reduces welfare when the service externality is not too strong. But as the externality becomes stronger, consumer losses increase because of the reductions in retail service. Thus, vertical integration will increase welfare at some point when the service externality becomes strong. There are three separate welfare effects of vertical integration. The reduction in retail diversity works to decrease welfare. On the other hand, eliminating the wholesale markup and internalizing the service externality both work to increase welfare.

In summary, a complete welfare evaluation of vertical integration requires an examination of the relevant non-price dimensions of competition. The problem is that such models become very cumbersome, even with convenient parameterizations and assumptions.

4. Uncertainty and information

In this section we discuss the vertical integration caused by the presence of uncertainty or private information.

4.1. Diversification

When all intermediate markets are competitive without imperfections such as rationing, firms are still subject to fluctuations in the prices which result from exogenous shifts in supply or demand at any point in the system. However, vertical integration cannot insulate the firm from such uncertainties. To illustrate this point, consider the simple model by Perry (1982). Consumers purchase the final good from retailers who in turn obtain the good from an intermediate market supplied by manufacturers. There is also an external net demand for the good from intermediate market. This can be thought of as foreign demand or supply of the manufactured good. Finally, manufacturers purchase the factor inputs necessary to produce the good from factor markets. Exogenous random events can impinge upon the final domestic market through shifts in consumer demand, upon the intermediate market through shifts in net foreign demand, or upon the factor market through shifts in the factor supply.

The manufacturers and retailers can vertically integrate in the sense of joint ownership, but the integrated firm cannot and will not want to avoid the price fluctuations in these markets. Even if the integrated firm refused to participate in the intermediate market, thereby synchronizing its manufacturing and retailing subsidiaries, it would still be directly subject to shifts in consumer demand or factor supply, and indirectly subject to shifts in net foreign demand in the intermediate market through its impact upon the factor and final prices.[15] Moreover, the integrated firm should optimally continue to participate in the intermediate market just as if its subsidiaries were independent firms. The price in the intermediate market is the relevant opportunity cost for each of the

[15]Bernhardt (1977) examines the extent to which forward integration and synchronization allows the manufacturer to reduce the variability of its demand. When the source of the fluctuations is consumer demand, integration cannot reduce demand variability for the manufacturer. However, when the source of the fluctuations is random purchasing behavior by retailers, integration can reduce demand variability.

subsidiaries [see Porter (1980)]. Thus, if the intermediate price is high, the manufacturing subsidiary should expand production while the retail subsidiary should contract domestic sales.[16]

Since no real production decisions are affected,[17] why then should vertical integration arise? Diversification of the firm's returns is the only possible incentive. Even this requires that capital markets be imperfect in that investors cannot adequately diversify against these fluctuations in their portfolios. When the source of the fluctuations is exogenous shifts in final demand, the returns of the manufacturers and retailers are positively correlated. This increases the fluctuation in profits and diversification would favor vertical disintegration. However, when the source of the fluctuations is exogenous shifts in foreign net demand in the intermediate market, the returns of the manufacturers and retailers are negatively correlated if the elasticity of final demand exceeds the elasticity of substitution in retail production. This would create a diversification incentive for vertical integration.[18]

4.2. Rationing and assurance of supply

The traditional business explanation for vertical integration is that firms want to assure their supply of inputs and their market for outputs [see Porter (1980)]. This typically means much more than mere avoidance of purely random fluctuations in the intermediate market.[19] Moreover, it must mean more than simply acquiring inputs at low prices or selling outputs at high prices. In particular, the notion of "assuring supplies or markets" entails the inability to obtain the quantity of inputs that the firm would wish to purchase at the prevailing input prices or the inability to sell the quantity of output that the firm would wish at the prevailing output prices. In other words, the market is not clearing at prevailing prices because of some imperfection.

[16]Of course, this assumes no economies from synchronization of production. See Subsection 7.4 for a vertical equilibrium model with such economies. Moreover, this assumes no costs of participation in an intermediate market subject to price fluctuations. Wu (1964) discusses such costs in suggesting that integration and synchronization is more profitable.

[17]Blair and Kaserman (1978) examine a model in which final demand is uncertain but that retailers must make their output decisions prior to the revelation of the uncertainty. Risk aversion will distort the output decision of the firms, inducing forward integration by a risk neutral manufacturer.

[18]When the source of the fluctuations is exogenous shifts in factor supply, it is not clear whether to expect the returns of the manufacturers and the retailers to be positively or negatively correlated. Levin (1981) finds that vertical integration into crude production reduced the variance in profits for oil refiners from 1948 to 1972. Mitchell (1976) argues that the cost of capital (S & P stock rating) is lower for integrated oil refiners.

[19]In 1981, Du Pont acquired Conoco stating that this would "reduce the exposure of the combined companies to fluctuations in the price of energy and hydrocarbons" [see Buzzell (1983)]. However, the fluctuations in oil prices during the 1970s were hardly random so that one expects that other presumed risks were being avoided by this merger.

Stigler (1951) pointed out that rationing induced by regulated prices would provide a powerful force for vertical integration. In a vertical equilibrium model by Green (1974), rationing results in an inflexible intermediate good price and provides an incentive for integration. But to understand the argument for "assuring supplies", one must account for rationing in a model with unregulated optimizing firms. Carlton (1979) has provided such a model.

In Carlton's model there are many consumers who demand a perishable good produced by the manufacturers. However, there is a retail stage and each consumer must seek the good only by presenting his demand function to a single retailer in each market period. Aggregate consumer demand is random, but all production and pricing decisions must be made before demand is observed. At the preset retail price, retailers attempt to satisfy the demands of their consumers, first from their own production of the good, i.e. vertical integration, and then by purchases in the wholesale market. But like consumers, retailers can frequent only one manufacturer in each market period. Thus, if a manufacturer has not produced a sufficient amount of the good prior to entering the market period, then some retailers will not be able to obtain enough of the good to satisfy all of their customers.

The production cost of the good, for either manufacturers or integrated retailers, is a constant c. However, demand uncertainty creates an additional cost from the risk that produced goods may go unsold. Thus, the equilibrium wholesale price set by the independent manufacturers, p_{int}, must exceed the marginal production cost c. Now, if a retailer were certain that at least one consumer would show up at his store, he could increase his profits by producing the good himself at the cost c rather than purchasing the good from a manufacturer. Thus, each retailer will partially integrate backward if the probability of no consumers showing up, Pr(0), is sufficiently low. This occurs when the cost of production, c, is less than the expected cost of having to enter the wholesale market to obtain one unit of the good, $[1 - \Pr(0)] \cdot p_{\text{int}}$. This incentive is reinforced by the probability that the retailer may also be rationed by the manufacturer and thus be forced to forgo the sale altogether.[20]

A vertical equilibrium will exist in which retailers partially integrate to satisfy the demand which will arise with high probability, and use the wholesale market to satisfy any greater demand which arises with low probability. This is the sense in which the model captures the notion of "assuring supplies". The traditional arguments did not account for why input supplies were unreliable, and appealed only to lost sales as the reason for seeking reliability through vertical integration.

[20]Carlton's model could also account for differing incentives to integrate. In particular, suppose the manufacturers sold an intermediate input which was employed by fabricators in differing consumer goods industries. The fabricators in some industries may have more certain demands, and thus have a greater incentive to integrate.

Not only is rationing endogenous in Carlton's model, but the insight behind partial integration is more appealing than the traditional arguments.

Vertical integration does not improve welfare in Carlton's model. As a result of integration, retailers now bear some risk of excess production. However, retailers are less efficient than manufacturers in bearing this risk. Since there are more retailers than manufacturers, the retailers have to produce a greater quantity of the good in order to satisfy the same fraction of consumers. Thus, vertical integration by the retailers, full or partial, will reduce the expected utility of the market equilibrium. Industry costs will be higher, prices will probably be higher, and consumers will incur a lower probability of obtaining the good.[21]

Carlton's model illustrates the incentive that rationing provides for vertical expansion. The retailers do not acquire manufacturers, but set up their own manufacturing subsidiaries. However, if we considered vertical merger in Carlton's model, we would discover a strong incentive for the manufacturers to acquire the retailers. The retailers owned by a given manufacturer would funnel their demands directly to him. In this way, the manufacturers could aggregate the demands of their retailers and the larger number of retailers would no longer interfere with efficient provision of the good to consumers by the smaller number of manufacturers. Exclusive dealing could also generate this outcome. It is also important to note that rationing in this model creates a strong incentive for horizontal merger among either the retailers or manufacturers.

4.3. Information and agency problems

Arrow (1975) posits a model in which vertical integration allows the acquisition of valuable private information. Downstream fabricators must make investment decisions prior to knowing the price of the intermediate input. Moreover, there is uncertainty about the intermediate price because the output of each upstream supplier is affected by the realization of a random variable. In general, the random variables will be correlated across suppliers. Backward integration enables a fabricator to learn the output of the acquired supplier prior to making its investment decision. Since the intermediate market is competitive, there is no rationing nor any incentive to internally transfer the input at other than the market price. Thus, backward integration enables the fabricator to obtain a better prediction of the intermediate price and make a more profitable investment decision.

Although the model does illustrate the acquisition of information, the upstream firms have no incentive to conceal their information. The upstream firms

[21] Carlton was unable to prove that the final price was always higher under partial integration. However, he was able to demonstrate that consumers would be worse off even if the price fell because of the lower probability of obtaining the good.

obtain knowledge about their supply prior to the investment decisions of the downstream firms, but Arrow assumes that the information is not revealed until aggregated into the equilibrium price. Since upstream firms are competitive, they cannot benefit from concealing their private information. Instead, they can benefit by simply revealing the information and improving the investment decisions of the downstream firms. Thus, vertical integration is not necessary. Downstream firms could buy the information from individual upstream firms or, better yet, an independent agency could collect and disseminate the aggregate supply information.

Crocker (1983) illustrates the acquisition of private information in a principal–agent model which avoids the degeneracy inherent in Arrow's model. The retailer (the agent) has private information about whether the final demand price is high or low. The manufacturer (the principal) then maximizes his expected profits by choosing a contract which specifies the quantity and the wholesale price as a function of the retailer's report. In order to induce the retailer to correctly reveal the final price, the contract offered by the manufacturer yields quasi-rents for the retailer when the final price is high, and results in less than the ex post optimal quantity when the final price is low. The retailer has no independent incentive to reveal its private information about the demand price since the strategic use of that information yields quasi-rents. Forward integration eliminates the inefficiency by enabling the upstream firm to directly observe the joint profits. The integrated firm can then induce the downstream subsidiary to reveal the correct price by rewarding the downstream subsidiary only when the optimal quantity for the reported price is consistent with the observed joint profits.

The inefficiency which arises from the profit-maximizing contract in the Crocker model is an example of the costs in using contractual exchange. However, there are no corresponding disadvantages of vertical integration. Vertical integration is posited as necessary to obtain the private information, but it creates no new problems such as attenuated incentives on the part of the downstream subsidiary to lower costs. Note also that this inefficiency could also be eliminated with more complex payment schemes designed into the contract [see Riordan (1984)].

Riordan and Sappington (1987) have a more complex model of vertical integration and private information. The model can be interpreted as having three vertical stages. Research and development determines the quality of the final good and the developer is endowed with private information about the unit cost of quality. The manufacturer of the final good has private information about the unit cost of production. Finally, the retailer merely sells the product at a known demand price. The retailer is the principal and the developer is the agent, but either can do the manufacturing and both are equally adept as such. Thus, the principal must decide whether to manufacture the product himself or let the

agent do the manufacturing. Whoever does the manufacturing observes the unit cost of production. When the cost realizations are positively correlated, the developer's incentive to overstate the cost of quality is accentuated when he is also manufacturing. As a result, the retailer's expected profit is lower when the developer manufactures the product than when the retailer himself does the manufacturing. Thus, the retailer would vertically integrate into manufacturing. Conversely, when the cost realizations are negatively correlated, the developer's incentive to overstate the cost of quality is dampened when he is also manufacturing. It may then be best for the retailer to have the developer do the manufacturing.

As in the models of Arrow and Crocker, vertical integration in the model of Riordan and Sappington enables the principal or the agent to observe valuable private information, the true unit cost of production (once the quality decision is observed). Moreover, like Arrow's model, the value of the information revealed by backward integration is that it may be correlated with the private information held by the agent, the cost of quality. However, unlike Arrow, there is no independent incentive to reveal the private information without vertical integration. And unlike Crocker, vertical integration by the principal does not eliminate the principal–agent problem and the associated inefficiencies.

It is not clear that vertical integration should be characterized as automatically revealing valuable information. The parties in the new integrated firm would often remain unchanged, and they would still have individual objectives apart from the firm's interest in profit maximization.[22] If private information is not revealed by vertical integration, a principal–agent problem remains within the firm. This raises the question of who should then be the principal within the integrated firm, the upstream subsidiary or downstream subsidiary. One might think that the acquiring firm, the principal, should be the one whose private information is the most valuable to the exchange and the most costly to obtain by internal incentives.

The literature also raises a broader conceptual problem concerning the nature of vertical integration. Should vertical integration mean an increase in the decision-maker's control over production and distribution or an increase in his information about the parameters of production and distribution? The two may seem inseparable. For example, in the model of Riordan and Sappington, backward integration into manufacturing enables the retailer to use the information about manufacturing costs to write a contract with the developer which improves its control over the quality of the product. However, it would seem more natural to define vertical integration as directly augmenting control, yet recognizing that control then facilitates revelation of private information. The

[22] Evans and Grossman (1983) take the extreme position that the multidivisional management of large vertically integrated companies via transfer prices does not reduce the information and agency problems encountered in arm's-length contracting.

new controls which come with vertical integration would enable the decision-maker to write new types of contracts. As such, backward integration in the model of Riordan and Sappington can be thought of as allowing the retailer to write a contract directly with the developer.

Williamson (1975) and others have argued that vertical integration enables new monitoring mechanisms which are more effective in obtaining private information (or in obtaining it sooner). For example, Alchian and Demsetz (1972) discuss the problem "shirking" in team production where the productivity of individual inputs is difficult to measure. They argue that firms arise because an owner-employer can subjectively evaluate the performance of individual workers and discipline individual team members. Similar arguments may be applicable to vertical integration which imparts control over the inputs of the neighboring stages. The controls of a disintegrated firm using contracts or market exchanges are weaker because they are restricted to measurable outputs rather than the inputs themselves. For example, Anderson and Schmittlein (1984) find that the difficulty in evaluating the performance of salesmen is a significant factor in accounting for the fact that electrical equipment manufacturers frequently employ a direct sales force rather than sell through independent sales representatives. This suggests that models of vertical integration and information need to focus upon the exact mechanisms by which integration enables information to be discovered, and how this differs from the mechanisms available in contracts or markets.

Even if vertical integration does not enable specific new monitoring mechanisms, it could reduce the agency problem in other ways. Vertical integration should increase the likelihood and duration of exchange between the two subsidiaries. This assurance of a continued relationship could facilitate the discovery or transfer of information. For similar reasons, Malmgren (1961) argues that vertical integration simply requires less information than using contracts or markets. By committing itself to internal exchange, the integrated firm need not incur the costs of following market developments and investigating alternative sources of inputs or outlets for its products. Such a commitment presumably incurs a corresponding cost of possibly missing advantageous external opportunities.

4.4. Prices versus quantities

If private information cannot be discovered by vertical integration, the existence of private information does have implications for the organization of an integrated firm. Consider the literature on "prices versus quantities" initiated by Weitzman (1974). Weitzman is concerned with the problem of a "planner" who wishes to maximize expected surplus in light of private information which is too costly to elicit from the parties through revelation incentives. But this is also the

problem facing the manager of an integrated firm with upstream and downstream subsidiaries. The upstream subsidiary has private information about a parameter that affects its costs, while the downstream subsidiary has private information about a parameter that affects its revenues.

If quantity is the internal decision variable, the manager sets the transfer quantity so as to maximize expected joint profit with respect to his distributions on the private information. Call this option 1, or the quantity option. On the other hand, if price is the internal decision variable, then there are two decentralized options. In option 2, the manager sets the price to maximize expected joint profit given that the downstream subsidiary chooses quantity to maximize its profits at that price. In option 3, the manager sets the price to maximize expected joint profit given that the upstream subsidiary chooses quantity to maximize its profits at that price. Since joint profit is the objective function for all three integrated options, they dominate decentralized bilateral monopoly outcomes. Thus, the question is which option generates the highest expected joint profits for the integrated firm. Assuming that the random variables characterizing private information affect only the marginal costs and revenues, Laffont (1976) shows that one of the pricing options dominates the quantity option, but that the quantity option then dominates the other pricing option. Option 1 dominates option 2 when the revenue function of the downstream subsidiary is more concave than the cost function of the upstream subsidiary is convex. This means that information about revenues is more valuable to the planner than information about the costs.

Flaherty (1981) discusses the vertical integration in this framework. In some sense, option 1 is "more" integrated than option 2 because it involves quantity controls rather than price controls. Thus, the tone, if not the intent, of Flaherty's discussion is that when the revenue and cost functions are such that option 1 is preferred to option 2, that we should expect firms to be vertically integrated. This ignores option 3 which Laffont has shown dominates both options 1 and 2 in these circumstances. But even so, this makes too fine of a distinction between what are essentially three forms of internal organization for a vertical integrated firm. However, this literature remains relevant to the theory of vertical integration because the incentive to integrate is enhanced when the internal organization of the firm is optimally configured to deal with undiscoverable private information.

5. Transaction cost economics

Coase (1937) argued that the key to understanding vertical integration would come not so much from understanding the vertical production relationships, but rather from understanding vertical exchange relationships. Like production,

exchange is costly. Moreover, vertical integration is simply one method of effectuating a bilateral exchange. Thus, an analysis of the costs of alternative methods of exchange is the required in order to understand vertical integration. Viewed from the transaction cost framework, the neoclassical focus on market imperfections has limited value because it ignores the costs of exchange. Since the primary alternative to vertical integration is contractual exchange, transaction cost economics examines the relative costs of contractual versus internal exchange.

Oliver E. Williamson is the leading proponent of transaction cost economics, and his many writings on the subject are summarized in his two books *Markets and Hierarchies* (1975) and *The Economic Institutions of Capitalism* (1985). I cannot do justice to the richness and complexity of transaction cost economics in this chapter. However, I will briefly summarize the theoretical and empirical analysis as it applies to vertical integration.

5.1. Asset specificity

Bilateral monopoly between a buyer and seller arises because gains from trade are enhanced by investments in assets which are specialized to their exchange. Williamson calls this "asset specificity". Asset specificity may arise from investments in (1) specific physical capital, (2) specific human capital, (3) site specific capital, (4) dedicated capital, or (5) brand name capital. Such transaction specific assets give rise to what Klein, Crawford and Alchian (1978) call "appropriable quasi-rents", which are the difference between the value of the asset in this use and the value in its next best use. When the environment is complex and uncertain, the transaction costs of negotiating and enforcing contracts make it prohibitively costly to write long-term contracts which specify all obligations under all contingencies. The resulting bilateral relationships fail to define the terms of performance for the parties in all or certain states of nature. In such states, either party may engage in "opportunistic" behavior, attempting to extract the quasi-rents of the other party by threatening to dissolve the relationship unless price concessions are forthcoming. Goldberg (1976) calls this the "hold-up" problem. Opportunistic behavior involves costs of haggling, and may result in the failure to maximize joint profits.[23]

Some of these problems can be mitigated by provisions in long-term contracts designed to govern the exchange. These would include requirements clauses, price indexing clauses, cost-plus pricing clauses, liquidated damages, and arbitration

[23] Fellner (1947) discusses bilateral monopoly, and Machlup and Taber (1960) point out the obvious incentive to integrate if the bilateral monopolists fail to maximize joint profits.

provisions.[24] However, when asset specificity is substantial, contractual governance over opportunism may become very costly. Internal organization of this exchange through vertical integration may then be a more efficient governance structure. Of course, the hierarchical structure of internal organization gives rise to other costs such as attentuated incentives and bureaucratic distortions. Thus, given the degree of asset specificity, the relative costs of governance dictate the choice between contractual exchange and vertical integration. If asset specificity is weak, contractual exchange is preferable; whereas if asset specificity is strong, vertical integration is preferable.

Williamson (1985) complements this analysis of relative governance costs with relative production costs. For example, as the supplier invests in assets more specific to this particular buyer, it loses economies of scale or scope because of its inability to make sales to other buyers. Thus, contractual exchange has an additional production cost advantage for a given degree of asset specificity. This extends the range of asset specificity over which contractual exchange is preferable.

Riordan and Williamson (1985) enhance Williamson's analysis by making the degree of asset specificity a choice variable for the buyer and seller. The degree of asset specificity and the quantity exchanged are both greater with vertical integration than with contractual exchange. Thus, the choice between vertical integration and contractual exchange involves a comparison of joint profits at the optimally chosen quantity and asset specificity. Riordan and Williamson also specify production cost advantages of contractual exchange and design benefits of vertical integration, both of which yield the expected effects.

Riordan and Williamson specify, rather than derive, the production and governance costs relationships. Asset specificity is treated like capital in a variable cost function. It reduces the total and marginal costs of production, and is available at a constant marginal cost itself. The governance cost functions for contractual exchange and vertical integration are chosen to generate the interrelationships postulated by the theory. These specifications allow them to focus immediately upon the comparative statics of the Williamson analysis. Moreover, arbitrary specifications may be unavoidable for certain nebulous governance costs. However, once asset specificity becomes a choice variable rather than an exogenous technological condition, it would be preferable to have a model which directly derives the benefits and costs of asset specificity. In so doing, one would also need to determine how to parameterize the specificity of assets.

The production cost or design advantages of asset specificity raise market issues beyond the pure transaction costs issues of contractual exchange. The

[24]Goldberg and Erickson (1987) discuss such provisions in the long-term contracts used in the petroleum coke industry. Joskow (1985) does the same for the long-term contracts between coal mines and utilities with coal-burning electric generating plants.

supplier is forgoing sales to other buyers in the input market by investments in a specialized input for this buyer. Similarly, the buyer's demand for the specialized input is derived from investments in the differentiation of its product in the final market. These observations raise issues of economies of scale and scope on the supplier side of the exchange and issues of product differentiation on the buyer side of the exchange. Thus, the market alternatives surrounding the buyer and seller become important in determining the choices of asset specificity. The exchange remains bilateral, but it is circumscribed by the surrounding markets. Production and demand considerations as well as transaction costs become intertwined in an equilibrium specificity of the assets.

Porter and Spence (1977) examine a simple model with design specificity and market alternatives. Standardization of intermediate inputs widens the market for any given supplier, and potentially allows the suppliers to achieve greater economies of scale. On the other hand, a specialized input may enhance the value to consumers of the buyer's final product. The specialized input increases revenues for an integrated buyer. But his average cost for the input will be higher than the market price because the fixed costs of producing the input cannot be spread over other buyers. Larger buyers will integrate and smaller buyers will employ the standardized input. Porter and Spence do not characterize the equilibrium price for the standardized input. But this price would determine a vertical equilibrium in that buyers above a certain size would find it optimal to integrate. Despite the fact that standardization may maximize welfare, buyers may vertically integrate because they fail to take into account their impact on increasing the price of the standardized input when they switch to a specialized input.

Porter and Spence do not discuss contracting. If contracting were possible, suppliers would have a choice about producing a standardized input at the market price or producing a specialized input for one buyer. This would then introduce a discrete choice on design specificity. The number of bilateral contracting situations would be dependent upon the market alternatives for both the supplier and the buyer. Moreover, these markets would affect negotiating positions of the parties and thus the nature of the contracts.

5.2. Evidence

Teece (1976) discusses vertical integration by petroleum refiners using the transaction cost framework outlined by Williamson (1975). Teece argues that although there are many crude oil producers, transaction problems arise from the high cost to refiners of a shortage or interruption in their flow of crude oil [see also McLean and Haigh (1954)]. Teece then finds that the major oil companies are integrated, producing between 50 and 90 percent of their crude requirements.

Similarly, the refiners have integrated into pipelines since they are the specialized investment linking refineries to the crude producers.

Levin (1981) challenges these findings with some regression results that yield no significant relationship between the profitability of oil companies and their degree of self-sufficiency in crude production. Levin also argues that since the divestiture of Standard Oil in 1911 there has been a wide variation in the degree of self-sufficiency among oil companies and no clear industry trend toward greater self-sufficiency. This conclusion also raises questions about the received business history that the offspring of Standard Oil who were left without crude properties subsequently reintegrated into crude production [see Johnson (1976)]. Since Standard Oil made investments based upon its monopoly and monopsony position, these investments probably differed from those that would have arisen under a competitive industry structure. Moreover, the size, location, and specificity of the investments were chosen within the context of the integrated structure of Standard Oil. Both of these considerations could have distorted post-divestiture markets, encouraging the vertical integration which occurred. If there is an ideal degree of self-sufficiency in crude oil, one might expect that refiners below that degree would be engaging more heavily in exploration than refiners above that degree.[25]

Teece (1976) also argues that forward integration by the refiners into marketing can be explained in terms of transaction cost considerations. However, the arguments involve retail diversity and service externalities similar to the neoclassical models discussed in Section 3. The forward integration of refiners into retailing has ebbed and flowed. Refiner owned and operated stations were predominant in the 1920s, but franchise stations emerged in the 1930s. However, refiner stations still constitute about 20 percent of the retail market.[26] If forward integration is an avenue of competition among refiners, we might expect more refiner operated stations during periods in which crude is plentiful and refiner capacity is underutilized. This suggests a model with vertical externalities that cannot be internalized by refiner price reductions.

These and other case studies [see deChazeau and Kahn (1959)] suggest that vertical integration by the major oil refiners is "tapered" both backward and

[25]It is interesting to note that Standard Oil produced less than 30 percent of its requirements of crude oil prior to the divestiture [see Johnson (1976)]. This fact seems inconsistent with both the monopsony incentive discussed by Perry (1978c) and the transaction cost arguments of Teece (1976).

[26]Refiner relationships with gasoline stations has always been a sensitive policy issue. In *Standard Oil Co. of California v. U.S.* (1949, 337 U.S. 293), the exclusive dealing contracts between refiners and their franchisees were held to be a violation of §3 of the Clayton Act. As a result, franchisees have the right to sell other brands. However, few do so. In *Exxon Corp. v. Governor of Maryland* (1978, 437 U.S. 117), the Supreme Court rejected a Commerce Clause challenge to a Maryland statute prohibiting refiners from operating retail service stations. The statute was enacted in response to evidence that refiner-owned stations had received preferential treatment during the 1973 oil embargo.

forward from the refining stage. The typical major refiner is substantially integrated into transportation of crude and wholesaling of refined products, but only partially integrated into crude production and retailing. Porter (1980) discusses tapered integration as a means of obtaining many of the benefits of vertical integration while avoiding many of the costs. In particular, Porter argues that partial integration allows the firm to adequately coordinate flows of intermediate inputs while providing both a competitive check on the integrated subsidiaries and a bargaining advantage with respect to the independent firms in these stages. In contradiction, Buzzell (1983) examined a large dataset of manufacturing firms, and concluded that tapered integration is less profitable than either no or full integration.

Stuckey (1983) examined vertical integration in the aluminum industry and argues that backward integration into bauxite and alumina by the refiners of aluminum can be explained by transaction cost considerations. Bilateral monopoly arises for three main reasons. First, efficient refining of alumina and aluminum requires large plants relative to the size of the market. Second, bauxite deposits are scattered around the world and refining involves a substantial reduction in volume. Thus, there are considerable transportation cost savings from locating the refining near the mining. Third, Stuckey points out that both bauxite and alumina are heterogeneous materials. For example, bauxite deposits differ in their alumina and silica content, while alumina may be either sandy or floury in composition. Other characteristics of bauxite affect the design of aluminum refineries.

Backward integration into automobile parts by the assemblers has also been explained by transaction cost considerations. Klein, Crawford and Alchian (1978) attribute the merger of General Motors and Fisher Body in 1928 to their close bilateral relationship and the inability to settle their disputes in the context of a long-term contract [see also Weiss (1987) for a discussion of the 1953 merger of Chrysler and Briggs Manufacturing]. Monteverde and Teece (1982a) use engineering effort in the design of components to measure transaction-specific skills, and find it highly significant in accounting for backward integration by GM and Ford. In related work, Monteverde and Teece (1982b) measure the quasi-rents in automobile components by calculating the cost of converting the tooling to its next best use. They then find that this measure is significantly related to the probability that assemblers own the tooling equipment. Since the equipment is employed by independent suppliers to produce the components, the arrangement is one of vertical quasi-integration. These explanations for backward integration in the automobile industry are in sharp contrast with the price discrimination explanation of Crandall discussed in Subsection 2.3. White (1971) presents a more complete picture of vertical integration by the U.S. automobile manufacturers. As with the petroleum industry, White concludes that tapered integration has been common for many components.

Using FTC data on mergers,[27] Spiller (1985) examines vertical mergers to test the transaction cost theory. Spiller estimates the gains from mergers using stock price information, and finds that they are negatively related to the distance between vertically related plants, a measure of transaction-specific assets. This provides some support for site specificity and vertical coordination as explanations of vertical mergers. Levy (1985) uses a distance measure as one of the explanatory variables for the ratio of value-added to sales for 69 manufacturing firms. It is intended to capture site specificity, but it is insignificant. However, Levy finds that the ratio of R&D expenditures to sales is significant, and argues that this supports the transaction cost theory since R&D is a proxy for specialized inputs. Armour and Teece (1980) argue that vertical integration increases the returns to R&D by facilitating communication between stages of production. They examine the U.S. petroleum industry from 1954 to 1975 and find a positive relationship between the number of stages in which firms participate and their expenditures on R&D.

Helfat and Teece (1987) argue that vertical integration enables the firm to better adapt to uncertainty and complexity, and that this should reduce the relative response of the firm's return to general economic uncertainties, measured by β in the CAPM model of financial markets. They examine 14 vertical mergers and find that the actual β of the merged firm is smaller (relative to a control group) than a predicted β derived from the premerger β's of the two merger partners. Weiss (1987) argues that the transaction specificity of two firms can be measured by observing that their stock returns are more highly correlated with each other than with other firms in the same industry. Using FTC data on mergers, Weiss finds that the returns of vertical merger partners prior to the merger were indeed more correlated. As Weiss points out, it remains difficult to explain why contractual alternatives became unsatisfactory at the point of merger.

Masten (1984) found a strong relationship between vertical integration by aerospace firms into the production of components and a measure of design specificity. The measure of design specificity was constructed from survey responses on whether the component could be adapted for use by other aerospace firms or by other industries. Anderson and Schmittlein (1984) examine the marketing decision by the producers of electrical components to use either a direct sales force or a system of manufacturer's representatives. They construct a transaction-specific human capital variable from survey responses, and find it to be a significant factor in explaining the use of direct sales forces. More generally, a direct sales force enables a firm to provide incentives such as security and promotion which may be better at eliciting specific human investments than the monetary rewards which are typical for representatives or franchisees. This is a

[27]The FTC has classified mergers involving assets in excess of 10 million dollars into horizontal, vertical, and conglomerate.

good example of vertical integration defined in terms of the firm's relationship to labor.

5.3. Cases from business history

Case studies in business history have been used by Williamson and others to illustrate various transactional reasons for forward integration. Chandler (1977) has employed such case studies to infer an intriguing link between mass production and vertical integration. Technological and organizational innovations during the nineteenth century permitted a smaller workforce to produce a higher rate of output. This was accomplished by a finer division of tasks to which specialized machinery was applied. Vertical integration was then required to synchronize the intermediate input flows between these stages of production within the firm. Technological or transactional reasons could account for the "economies of synchronization" which induce this integration within the production stage.[28] However, Chandler extends the argument to explain forward integration beyond production into wholesaling and retailing. Forward integration arose from the inability of traditional marketing channels to respond to the pressure created by the greatly increased volume of goods from manufacturers taking advantage of the economies of scale at the production stage.

Chandler's argument has its greatest appeal for the cases involving perishables. Swift developed and owned a fleet of refrigerated railroad cars and wholesale houses in major cities in order to distribute beef dressed in its midwestern meat-packing plants. This system replaced the delivery of beef on-the-hoof to eastern markets. Similarly, United Fruit owned banana plantations, a fleet of refrigerated steamships, and wholesale houses in major cities for ripening the bananas [see Porter and Livesay (1971) and Chandler (1977)]. These cases have been cited as illustrations of transaction-specific investments where failure to synchronize the flows imposes large costs on one party or another. However, the assets at each stage of these industries were specific to the producers primarily because the producers were monopolists. Assets specialized for a particular use become effectively transaction-specific when a neighboring stage is monopolized. Thus, integration into specialized transportation and wholesaling may not have been necessary if there had been more producers. Independent transportation and wholesaling companies could have switched producers in response to opportunism by one.

[28] McLean and Haigh (1954) focused upon this theme is explaining the backward integration by oil refiners into crude production. Unit costs increase rapidly with refinery operations below capacity, while the costs of maintaining inventories of crude are non-trivial. They argue that integration insures the optimal flow. Of course, long-term contracts may achieve similar results. In Goldberg and Erickson (1987), the costs of storing hazardous petroleum coke was a key consideration in long-term contracts for its disposal by oil refiners.

The industries which Chandler cites for the pure application of his coordination argument are tobacco, matches, grain milling, canning, soap, and photography. The manufacturers in these industries set up sales offices and warehouses in major cities to distribute their products. These offices often sold directly to large retailers. The manufacturers also initiated product differentiation by the advertising of new branded products. Thus, the traditional role of merchant wholesalers was substantially diminished. Chandler argues that forward integration and the switch to branded rather than generic products was a consequence of the need to market the greatly increased volume of output from the new continuous-process technologies employed in manufacturing. An alternative view is that branding was a marketing innovation independent of the technological innovations. Under this view, forward integration by manufacturers could have been the efficient method of distribution for the new branded products, or a temporary measure to deal with the sluggish response by the traditional distribution channels.

The experience in these industries could be also be understood as a tradeoff between economies of scale and scope. With indivisibilities in wholesaling, a substantial volume of sales would be necessary to justify setting up sales offices, even if such offices had coordination and promotional advantages. Otherwise, the economies of scope in wholesaling different products to the same retailers would dominate. In discussing the tobacco industry, Porter and Livesay (1971) point out that the manufacturers continued to use independent jobbers for distribution to small retailers in urban markets and to rural retailers. These jobbers could capture economies of scope by distributing a range of grocery and drug products.

Chandler also discusses forward integration into retailing by the manufacturers of consumer durables. These products required specialized services such as information, repair, and credit. He argues that existing distribution channels were unable or unwilling to provide these services. For example, Singer and the other major sewing machine companies set up branch offices for demonstrations, sales, and repair. Similarly, McCormick sold farm machinery through exclusive franchise dealers and had regional offices to handle repair and supervise advertising and credit by the dealers. Although Chandler discusses the inventory problems that arise with independent wholesalers and retailers, the synchronization of flows is not the most natural explanation for this forward integration. Incentive and externality problems seem more likely explanations. The incentive problem concerns the best compensation arrangement for generating sales and service effort. Sales and service may be done by independent distributors, by franchised distributors, or by employees. The externality problem concerns the spillovers and resulting incentives to free-ride in the provision of non-appropriable presale services.[29]

[29] Vertical quasi-integration and vertical controls are often employed to deal with incentive and externality problems in retailing. For example, see Mathewson and Winter (1984) and Perry and Porter (1986).

Williamson has used these cases as an example of forward integration motivated by human asset specificity. The argument is that manufacturers require sales and service people specifically trained to handle the complex differentiated products. This may have been true when these products were originally introduced. However, with growth and maturity, these and similar industries have attained a higher sales volume, a stable set of producers, and some standardization. These developments have given rise to independent service firms, most notably in the automobile industry. This suggests that human asset specificity may be more a consequence of product novelty rather than product complexity or differentiation.

The overwhelming aspect of the cases from American industry prior to 1900 is that the industries were built around new products as well as new technologies. Thus, to fully assess of Chandler's arguments, we need the entire case histories rather than case studies at the particular point in time when these new products were introduced.

6. Incomplete contracts

In this section we discuss the recent theoretical literature on "incomplete contracts" and the relevance of this work for understanding vertical integration. To a large extent, this literature is a theoretical response to transaction cost economics where the inability to write complete contracts allows opportunistic ex post bargaining. However, the models have no explicit transactions costs, nor any costs associated with the governance of a long-term contractual relationship. Rather, the losses arise from inefficiencies in the ex ante investments or the ex post exchange.

6.1. Contract law

An incomplete contract has been defined by economists as a contract that either fails to specify performance obligations for the parties in all states of nature, or fails to specify the nature of the performance itself [see the survey by Hart and Holmstrom (1986)]. The first case occurs because it is costly to enumerate the future states of nature or costly to agree about the performance obligations in a given state. The second case occurs because it is costly to agree on or specify clearly the performance obligations, irrespective of the state of nature. In addition, either case can also occur if it is impossible for a third party such as a court to verify the occurrence of a state of nature or identify a performance obligation.

When a contract is incomplete, economists may implicitly assume that the contract is not longer relevant for determining the nature of the exchange in the

first case or for determining some dimension of the exchange in the second case. If so, the contract fails to circumscribe the bargaining and reduce opportunism. This favors vertical integration as the preferable form of exchange. A summary of contract law will demonstrate that this is a narrow and inaccurate conception of contracting. In the process, we will clarify some issues which are relevant for theoretical work.

The term "incomplete contract" has no independent meaning in contract law. However, the economic notions of incompleteness do have legal analogies. A contract which specifies performance obligations only for certain states of nature would be a contract "with conditions precedent". A "condition precedent" is an event, other than the lapse of time, which must occur before a performance obligation arises.[30] A contract with conditions precedent will then have no performance obligations in the unspecified states of the world. For example, contracts between a general contractor and a subcontractor are typically contingent upon the general contractor winning his bid. If exchange is to occur in the unspecified states, a new contract must be negotiated.

Contracts which fail to specify the nature of performance are called "indefinite contracts". Contracts may be so indefinite that they would have no legal force. In early contract law, any indefiniteness about a material term such as quantity, quality, price, delivery, or payment would invalidate the contract. Williamson (1979) refers to this as "classical" contracting. However, modern contract law governed by the *Uniform Commercial Code*[31] (*UCC*) will attempt to "fill the gaps" of an indefinite contract, and invalidate it only if a court could not fashion a remedy upon breach.[32] In supplying the unspecified terms, called "open" terms, the courts will examine the course of dealing by the parties on prior contracts, the course of performance by the parties on this contract, or the customary practice in the trade (*UCC* 1-205). A court will even infer a "reasonable" price for a contract which fails to mention price [*UCC* 2-305 (1)(a)]. Note that the reasonable price need not be the market price.

The parties may purposely leave the performance obligations indefinite by "agreeing to agree" at a future time before performance is due. But, like indefinite contracts, a court will substitute a "reasonable" performance if the parties fail to agree. The parties may also leave the performance obligations indefinite by allowing one party to decide certain terms at a future time. The

[30]See Calamari and Perillo (1977, ch. 11). The party to whom the performance obligation is due must prove the occurrence of the condition.

[31]The *UCC* is a codification of contract law and has been adopted with minor variations in every state but Louisiana.

[32]*UCC* 2-204(3) states that "even if one or more terms are left open a contract for sale does not fail for indefiniteness if the parties have intended to make a contract and there is a reasonably certain basis for giving an appropriate remedy". Of course, the cost of enforcing a contract with more gaps would increase for the non-breaching party.

most common example is a long-term requirements contract in which either the buyer (or seller) can choose the quantity exchanged in each future period within some bounds [*UCC* 2-306 (1)]. Indeed, the *UCC* even allows contracts in which either party can set the price for a future exchange [*UCC* 2-305 (2)]. Open terms of performance assigned to one party would seem to be an invitation to opportunism, but the *UCC* requires that these terms be set in "good faith". If the open terms are not set in good faith, a court will again substitute "reasonable" performance.

This discussion is not intended to suggest that opportunistic bargaining cannot arise under a contract with indefinite or open terms of performance. Rather, the point is that such terms do not nullify performance obligations as if a contract never existed. Thus, unlike contracts with conditions precedent, indefinite contracts create performance obligations for each party in all states of nature. If a dispute arises as to the nature of the performance, its resolution will be negotiated in light of the rights and remedies under contract law. Of course, the legal positions of the parties may be uncertain, and there are costs of enforcing those rights and remedies. But the bargaining position on an indefinite contract would generally be different from the bargaining position on a contract with a condition precedent.[33] Indeed, one should expect that even indefinite contracts are intended to circumscribe the range of bargaining outcomes and thereby reduce the potential for subsequent opportunism.

Even if a contract has no conditions precedent and no open terms of performance, renegotiation can occur in response to contingencies. Parties to a contract have the right to not perform their contractual obligations, i.e. breach the contract. But the breaching party is then subject to money damages sufficient to compensate the non-breaching party for his expected value under the contract. Breach and the payment of damages enables the parties to avoid an exchange which has become inefficient under realized states of nature not specified as conditions precedent. Even if the exchange remains efficient, one party may incur a loss under the contract for some states of nature. Since a contract can be renegotiated and modified even though the performance obligations of only one party are reduced (*UCC* 2-209), the right to breach could also trigger renegotiation in this case.

Shavell (1984) points out that money damages for breach and renegotiation are legal substitutes for conditions precedent in contracts. This is particularly true if the conditions precedent would not be verifiable by a third party such as a court or arbitrator, and thus are not contractable. Since all that can be observed is non-performance, a court must fashion a remedy which is independent of the

[33]There could be dispute about the occurrence of an event which would trigger performance obligations in a contract with conditions precedent. If so, the bargaining positions would be based upon a combination of the legal and economic positions.

state of the world that triggered the breach. Renegotiation would then prevent the breach of an efficient contract (if there are no costs of enforcement).

When there are legal costs of enforcing the payment of damages, an opportunistic party can extract some quasi-rents from the other party by threatening to breach even if it would not incur a loss on the contract. Renegotiation to reach a performance or monetary settlement may be better than going to court for the non-breaching party. Thus, renegotiation can occur in the context of a complete contract so that performance differs from that specified in the initial contract. However, the bargaining between the parties over the new performance obligation is still circumscribed by the legal rights and remedies under the initial contract.

In summary, a contract alters the ex post bargaining positions of the parties when there are no explicit conditions precedent to performance. The economic positions of the parties in the realized state of nature remain relevant in that they define the gains from exchange. However, the contract performance specified by the parties or by the law alters the positions of the parties if exchange fails to occur. Unless the costs of enforcement are very large, the remedies for breach of the contract replace the pure economic positions of the parties as the threat point for ex post bargaining.

6.2. Models of incomplete contracts

The incentive for pre-performance investments is a crucial issue in evaluating vertical integration relative to contractual exchange. Such investments widen the gains from trade by increasing the value of the exchange to the buyer, v, or decreasing the costs of production to the seller, c.

Tirole (1986) discusses the pre-performance investment problem. The seller's costs of performance in the second period decline with an investment in the first period and depend upon a random variable which he learns at the beginning of the second period. The value of the exchange cannot be enhanced by investments on the part of the buyer, and thus it depends only upon a random variable which he learns at the beginning of the second period.

Vertical integration might be thought of as the symmetric information complete contract in which the seller chooses the level of investment which maximizes the discounted expected net value of exchange in the second period minus the cost of investment in the first period. Since only the ex post efficient exchanges are consummated in the second period, performance in various states of nature need not be specified in the first period. Williamson (1975) calls such a process "adaptive sequential decision making". In this sense, vertical integration avoids advance specification of any contingencies.

Tirole then focuses upon the short-term contractual alternative. At the beginning of the first period, the buyer and seller cannot contract on the price in the second period. However, if the investment is observable and verifiable, they can contract on the level of investment and its financing. The price is then determined by bargaining in the second period subject to the private information of the parties.[34] Thus, the contract is incomplete in the sense that it does not specify price or performance in the second period. There is not even an agreement to agree since the bargaining process need not result in an exchange (and surely not if costs c exceed the value v). The buyer can only bargain for the investment by the seller in the first period.

Tirole finds that the investment contracted for need not be less than the optimal level of investment. The buyer may bargain for overinvestment by the seller because the resulting cost reduction can soften the seller's behavior in the second period bargaining process.[35] Thus, if vertical integration is interpreted as the symmetric information complete contract, the investment under the short-term contract is non-optimal and may actually exceed the investment under vertical integration.

It is surely inappropriate to interpret Tirole's symmetric information complete contract as vertical integration. As such, vertical integration allows full use of all the information of both the buyer and the seller in the second period. Instead, vertical integration might be defined as one party choosing both the level of investment in the first period, and whether exchange takes place in the second period. However, in deciding whether exchange takes place, the integrating party would not have access to the other party's private information. Vertical integration would only enable him to structure incentives for obtaining private information. This would capture the notion of a long-term relationship without assuming the acquisition of information. Since some probability of an inefficient exchange would then exist, the investment could be non-optimal under vertical integration.

Tirole does not examine long-term contracts for which the price in the second period is contracted in the first period. This limits what can be inferred about vertical integration because long-term contracts are the best alternative to vertical integration. Hart and Moore (1988) focus upon such a contract and allow renegotiation in light of the right to breach and pay damages. At the beginning of the first period, the buyer and seller agree to a contract price for exchange in the second period. The buyer and seller can then take actions such as investments in the first period which will respectively increase the value v or reduce the cost c of

[34] If investment is unobservable and thus non-contractable, a weak assumption on the bargaining process implies that the seller will choose a level of investment less than the optimal level chosen under vertical integration. The bargaining solution prevents the seller from reaping the full reward of his investment activity.

[35] With symmetric information, efficient bargaining in the second period enables the buyer to contract for the optimal investment in the first period.

the exchange. However, these investments are not observable and thus cannot be contracted for themselves. As a result, Hart and Moore do not examine vertical integration which would have an impact upon these investments.[36]

The state of nature is realized at the beginning of the second period revealing the value and the cost to both the buyer and seller. Renegotiation then occurs to prevent one party from breaching if the exchange is efficient, i.e. $v > c$. The compensation to the party who would breach under the contract price is increased to make him just willing to make the exchange. The other party than absorbs the surplus $v - c$ of the exchange. Note that this particular bargaining solution is only triggered by the certainty of breach. As such, it implicitly assumes that specific enforcement of the contract is costless. Thus, there is no real opportunistic bargaining in the second period. However, this division of surplus in the second period still produces too little investment in the first period.[37]

Hart and Moore have designed their bargaining model around the existence of a long-term contract in which the contractual obligations can be costlessly enforced. Thus, the model does not capture all the bargaining issues inherent in renegotiation of a long-term contract. If there are costs of enforcement, a bargaining game of renegotiation could involve opportunistic rent extraction up to the amount of these costs. The ability of the non-breaching party to obtain specific performance upon incurring these legal costs prevents any greater rent extraction. However, money damages rather than specific performance is the typical remedy in commercial contracts not involving real estate. If the damage measure is less than the ex post net value to the non-breaching party, the potential rent extraction for the breaching party could be greater than the legal costs of enforcement.[38] Thus, even a long-term contract does not eliminate opportunistic renegotiation. However, it does place constraints upon bargaining outcomes which are different than those which exist when there is no contract.

The model of Tirole suggests directions toward defining vertical integration, while the model of Hart and Moore suggests directions toward defining long-term contracts. Grossman and Hart (1986) examine a related model in order to

[36] In a similar model, Riordan (1986) defines a contract which maximizes expected joint surplus. This contract involves a delegation scheme which may resemble how an integrated firm would function in this setting.

[37] If there is no probability that the exchange will be inefficient, a two-price long-term contract can be constructed which will generate the optimal investments. Hart and Moore do not compare the levels of investment with and without the long-term contract.

[38] Under a pure *expectation measure* of damages, the non-breaching party would receive the ex ante expected value of the bargain, $E(v) - p - I_b$ for the buyer or $p - E(c) - I_s$ for the seller. Of course, the actual damage awards by a court could deviate from these measures in that realized values or costs might be taken into account. Under a pure *reliance measure* of damages, the non-breaching party would simply receive his investment expenses. See Shavell (1984) for a discussion of damage measures.

explicitly discuss vertical integration. The two firms to an exchange can make pre-performance investments, but vertical integration is not defined in terms of control over these investments. Vertical integration is defined as control of subsequent production decisions.

The manufacturer must make his investment in the first period prior to a production decision about the quality or quantity in the second period. Similarly, the retailer must make his investment in the first period prior to a production decision about sales effort or service in the second period. Product-specific advertising is an example of an investment that might be undertaken by either the manufacturer or the retailer. There is no uncertainty or private information. Thus, without the sequencing of decisions, the manufacturer and retailer should be able to reach an efficient agreement in the first period on both investments and production decisions. However, Grossman and Hart assume that the production decisions are sufficiently complex that this complete contract is not possible.

Unlike the previous models, Grossman and Hart circumscribe the bargaining in the second period by introducing "residual rights of control" over the production decisions. In the absence of a contract in the second period, the manufacturer and the retailer each make their respective production decisions. Vertical integration is then defined as transferring these residual rights of control over the production decision exclusively to either the manufacturer or the retailer. Vertical integration does not transfer any rights to control the investments, and thus cannot directly solve the problem of inefficient investments. For this reason, their definition might be better interpreted as a contract which assigns vertical control to either the manufacturer or the retailer.

Without vertical control, the residual rights prevail. The Nash equilibrium in production decisions can then be defined for the given investments. This will be an inefficient solution in that it will not maximize joint profits. As a result, Grossman and Hart allow the production decisions to be contractable after the investments are made. The Nash equilibrium in the production decisions then becomes the threat point for a Nash bargaining outcome in those decisions. The investments in the first period are then chosen non-cooperatively in a Nash equilibrium with full recognition of the Nash bargaining outcome in the second period. Since neither firm internalizes the full incremental benefits in the second period from their investments in the first period, both firms underinvest.

With vertical control, the production decisions are vested solely with either the retailer or the manufacturer. Given the investments, the firm in control chooses both production decisions, so as to maximize his benefit as if no recontracting would occur. This choice again serves as the threat point for a Nash bargaining solution in the second period. Again, the Nash equilibrium for investments in the first period can be defined with full recognition of the Nash bargaining outcome in the second period. In general, the investments will be inefficient for the same reason as in the case of of no vertical control.

Grossman and Hart then compare these three alternatives. No vertical control is preferable when the production decision of each firm has only a small impact upon the benefits of the other firm. Vertical control by the manufacturer would result in a retail production decision which ignores the benefits to the retailer, while generating no offsetting benefits for the manufacturer. This can then produce a very inefficient investment by the retailer. In this sense, the model captures the notion of "opportunism" by the manufacturer. On the other hand, vertical control will dominate in certain circumstances. If the benefits of the retailer are relatively insensitive to either production decision, while the benefits of the manufacturer are sensitive to both production decisions, then vertical control by the manufacturer is preferable. Vertical control by the manufacturer in the second period insures a more efficient investment by the manufacturer in the first period. Either of the other alternatives could produce a very inefficient investment by the manufacturer. The analysis is identical when the roles of the retailer and the manufacturer are reversed.[39]

The model of Grossman and Hart has no long-term contracts with performance obligations. The investments are never contractable, and the production decisions are only contractable in the second period as a short-term contract. However, there is a long-term contract which allocates the rights to control the production decisions if negotiation of a short-term contract in the second period should fail. Such long-term contracts are indefinite contracts in which the terms of performance are left to be decided by the assigned party. But since the indefinite terms are chosen on the basis of self-interest rather than in good faith, these contracts have the flavor of ownership as well as control. This is the sense in which the model addresses vertical integration.

There are two reasons why this is an unusual definition of vertical integration. First, the production choices of the integrated firm merely serve as the threat point for the Nash bargaining outcome. As a result, vertical integration does not allow the integrating firm to internalize the full benefits of its choice of production variables. Moreover, this resembles renegotiation of an initial contract concerning vertical control, rather than integration. Second, the investment decisions cannot be directly affected by vertical integration. Vertical integration need not provide the direct control and information necessary to choose the optimal investments. This contrasts sharply with the transaction cost literature which views vertical integration as a means to alleviate opportunism and the resulting underinvestment. Even a neoclassical model of vertical integration would have the integrating firm exercise some control over the investments of the acquired firm in the neighboring stage. Indeed, vertical integration could occur by

[39]Note that these results are predicated upon an assumption that the investments of both firms are equally important to the outcome. If not, then vertical control by the firm whose investment is most important would generally be preferable.

having a third party owning the upstream and downstream subsidiaries but involving itself only in the investment decisions of these subsidiaries. The subsidiaries would make production decisions, dealing with each other at arm's length but subject to corporate rather than market incentives.[40]

7. Vertical equilibrium

When we examine the patterns of vertical integration, we not only find differences across industries, but we also find considerable differences among firms within an industry. This section discusses some models which attempt to account for these observations. In so doing, we introduce the notion of a "vertical equilibrium".

7.1. Vertical equilibrium

Coase (1937) outlined the equilibrium condition for vertical integration by an individual firm. Transactions are integrated when the internal cost of exchange is less than the external cost of exchange, either the cost of using markets or the cost of contracting. For Coase, this first-order condition for minimizing the costs of exchange defines the optimal size of the firm.[41] However, it seems more natural to interpret the Coase condition as the first-order condition on the extent of vertical integration of the firm, with the size of the firm remaining indeterminant.

The costs and benefits of vertical integration by a given firm would in general depend upon the extent of vertical integration by other firms in the industry. This suggests that firms in the same industry need not be equally integrated into the various stages of production. A "vertical equilibrium" would define a pattern of integration in the industry such that no firm would alter its choice of the stages in which it operates. As a result, firms may sort themselves, either by integrating or by specializing in certain stages of production. When combined with the existing models of "horizontal equilibrium" (competitive, oligopolistic, or collusive), a vertical equilibrium could complete the conceptual picture of the industry.

7.2. The Young model

In the *The Wealth of Nations* (Book 1, chs. 1–3), Adam Smith (1776) introduced the concept of "division of labor" and how it is necessarily limited by the "extent

[40] This would be a multi-divisional corporate structure of the integrated firm. See Williamson (1975, 1985).

[41] The Coase tradition, viewing the theory of the firm in terms of transaction costs, is in contrast with the neoclassical tradition of defining the firm in terms of a production function.

of the market". The insight of Smith is that a greater quantity of production will enable and induce a greater refinement of tasks within the firm. The growth in demand for a firm's product stimulates specialization by workers in that each worker performs fewer of the tasks required to produce the good. This reinvestment and specialization of tasks allows a greater quantity to be produced in a shorter time with less labor. Each worker becomes more skilled at this task and avoids losing time from switching tasks. Moreover, specialization stimulates each worker to devise better methods of performing his task. These new methods generally involve the application of machinery. In these ways, Smith argues that the division of labor is an important source of industrial progress.

Young (1928) agrees that growth in demand generates a refinement and specialization of tasks within the firm. However, he points out that the division of labor also applies to the development and production of machinery for the refined tasks. Thus, he argues that industrial progress arises both from applying specialized machinery to the refined tasks within the firm and from refining and specializing the tasks in the production of the machinery itself.

Vassilakis (1986) attempts to capture the Young story with a vertical equilibrium model. The final good can be produced from any one of a continuum of intermediate goods indexed by $v > 0$. Production "roundaboutness" is represented by higher v, and has implications for both labor cost and productivity. More roundabout production involves higher fixed costs of labor, $\gamma(v)$, where $\gamma'(v) > 0$, but lower marginal costs of labor, $c(v)$, where $c'(v) < 0$. The fixed costs increase as labor is employed to build new equipment, but the marginal costs of producing the intermediate good are lower because the new equipment requires less labor to operate. In addition, an intermediate good with a higher index may be more effective in producing the final good. If $\delta(v)$ is the number of units of the intermediate good v required to produce one unit of the final good, fewer units are required when production is more roundabout, i.e. $\delta'(v) \leq 0$.

The model has a fixed number of firms, each endowed with a unit of labor. The "specialist firms" are downstream firms which produce the final good from purchased units of whatever intermediate good is offered. Integrated firms produce the intermediate good both for sale to the specialist downstream firms and for production of the final good by their own downstream subsidiaries. Given final demand, general equilibrium in the labor and intermediate markets generates an equilibrium degree of production roundaboutness. This determines the boundary between the two stages of production. When the equilibrium wage rate is high relative to the equilibrium price of the intermediate good, production becomes more roundabout as labor is employed to create new equipment and thereby reduce the labor operating costs of producing the final good. The equilibrium number of integrated firms is then determined by the tradeoff between economies of scale in the production of the intermediate good and imperfectly competitive behavior by the integrated firms in the market for the

intermediate good. If there were only a few integrated firms, the intermediate good price would be high making these firms very profitable. Entry would then occur by downstream firms integrating into the production of the intermediate good, both to supply their own downstream production (avoiding the oligopoly markup) and to profit from sales to the remaining specialist downstream firms. Thus, the sorting of firms is determined simply by the relative profitability of being integrated or being specialized. With a larger economy (a larger number of firms/agents), the fixed labor costs and imperfect competition become less relevant. Thus, the degree of roundaboutness and the ratio of downstream to integrated firms both increase.

The model of Vassilakis is very intriguing, but it has some conceptual limitations. First, it only determines the boundary between two stages of production. As such, it does not have the full flavor of the Young story in which industrial progress results in *more* stages of production, i.e. a refinement of the producers' goods industries. This problem could be remedied by having a series of these models generating differing types of downstream firms using different intermediate goods. But this raises a second issue. A downstream firm which uses an intermediate product from a more roundabout production process is not more integrated itself because it is not really performing any of the intermediate stages. Thus, even if there were different types of downstream firms, this could not naturally be interpreted as differing degrees of integration. There would remain only one type of integrated firm. Finally, there would still be no specialized upstream firms, only integrated firms and specialized downstream firms. Once a firm is producing the intermediate good, there is no disadvantage of then producing the final good.

7.3. The Stigler model

The classic paper by Stigler (1951) not only provided the first review of the literature on vertical integration, but also introduced a novel theory of vertical disintegration in growing industries. Applying Adam Smith's theorem that the division of labor is limited by the extent of the market, Stigler first argued that infant industries would be composed of integrated firms because the level of production at any one stage is too small to support specialized firms and intermediate markets. However, as demand grows for the final good of the industry, stages subject to increasing returns would be spun off. At first there would be monopoly in these stages, but with continued industry growth, oligopoly and then competition would arise. Specialization drives vertical disintegration as the industry matures. Conversely, vertical integration would occur in declining industries. Stigler clearly envisioned a vertical equilibrium intertwined with the

horizontal equilibria at each stage of production. Moreover, the vertical equilibrium is derived from the production technology at the two stages, and not the exchange technology between them.

The Stigler model is concerned with the impact of industry growth upon the industry equilibrium. In contrast, it has been commonly argued that individual firms integrate when they have grown sufficiently at their primary stage in order to capture the scale economies at neighboring stages. This scenario implicitly requires that the diseconomies of scale at the primary stage be relatively minor, and that there be at least some economies of integration between the stages. To the extent that firm growth results from industry growth, the prediction of integration is in sharp contrast with the Stigler model. However, our interest in the Stigler model stems from the fact that it explicitly attempts to consider the interactions with other firms.

The difficulty with the Stigler model occurs from trying to capture the notion of specialization. If specialization means only that the production process of the upstream stage is subject to economies of scale, then the vertical equilibrium will degenerate irrespective of the size of the industry. As soon as demand has grown to the point that two downstream firms are more efficient than one, the upstream stage would be spun off. Vertical disintegration would prevail thereafter. Economies of synchronization could prevent this degeneracy, but Stigler assumes none.

This does not mean that the upstream stage could immediately set the monopoly price because downstream firms still have the option to reintegrate. Thus, the upstream monopolist would set a limit price. Further growth in demand would generally cause the input price to decline. First, demand could shift outward such that the monopoly price fell below the limit price. Second, even if the monopoly price remained above the limit price, the upstream stage could become so profitable as to induce entry of a second firm. The duopolists could charge the limit price or some equilibrium price less than the limit price. Thus, the only issue raised by growth in demand is the exact nature of the free entry equilibrium in the upstream stage, i.e. the price and number of firms.

An alternative view of specialization is that there are economies from doing a limited set of activities, rather than economies of scale from simply doing a lot of any one activity. These notions are inversely related if the firm is fixed in overall size, but they are distinct when we are explicitly discussing the size and boundaries of the firm. Specialization would mean diseconomies of scope across vertically related production processes.[42] Thus, both economies of scale and diseconomies of vertical scope would be necessary to generate large specialized firms as suggested by Adam Smith.

[42] Baumol, Panzar and Willig (1982) have discussed the difference and relationship between economies of scale and scope in the context of production processes which are not vertically related. With both types of economies, large diversified firms would arise.

Perry and Groff (1982) have constructed a model which resembles the verbal model of Stigler. The model avoids the degeneracy inherent in defining specialization as economies of scale. However, it does not possess true diseconomies of vertical scope. In particular, it is not more costly for a given firm to operate in both stages rather than operate in each stage separately. However, the notion of specialization is captured in a related, though ad hoc, fashion. Each firm is endowed with a separate cost function for each stage, but firms which are more efficient (lower costs at any output) in upstream production are less efficient in downstream production. In upstream production, there is a constant marginal cost for all firms, but the more efficient firms have a lower fixed cost. In downstream production, firms have linear rising marginal costs (and no fixed costs), but the marginal costs of more efficient firms rise more slowly. Thus, as posited by Stigler, there are economies of scale in the upstream stage and diseconomies of scale in the downstream stage. But since the cost function of the integrated firm is the sum of the upstream and downstream costs, there are no diseconomies of vertical scope for a given firm.

The cost functions are defined so that the integrated firms all have the same minimum average cost. But even though all firms are equally efficient in integrated production, many firms will specialize into either upstream or downstream production based upon their relative advantage. The firms with low fixed costs become specialist firms upstream, while the firms with flatter marginal costs become specialist firms downstream. This process is driven by imperfect competition at the upstream stage. The oligopoly equilibrium among the upstream firms generates a price above marginal cost. Firms with high fixed costs in upstream production pay this price for the input and specialize in downstream production. But firms with moderate fixed costs remain integrated in order to avoid the upstream markup. Given a linear final demand, a vertical equilibrium can be defined in which firms sort themselves into upstream, integrated, or downstream firms by the profitability of each alternative.

It is not surprising that firms would sort themselves according to their relative cost advantages. But the goal is to use this simple vertical equilibrium to examine Stigler's hypothesis that growing industries will disintegrate. Perry and Groff computationally examine the comparative statics of this vertical equilibrium for shifts in the demand parameters. Indeed, the Stigler hypothesis is confirmed for one very plausible set of circumstances. If demand shifts outward horizontally around a constant intercept (by decreasing the slope parameter) and if the total number of firms increases in the same proportion, then the equilibrium integrated sector decreases so that proportionately more firms specialize in either upstream or downstream production. However, the Stigler hypothesis is not confirmed when the number of firms does not increase with demand.

There is some empirical evidence in support of the Stigler model. Levy (1984) examined Census data for 38 industries from 1963, 1967, and 1972. He found

that demand growth, measured as the ratio of current to past sales, was significantly and positively related to the degree of vertical integration, measured as the ratio of value-added to sales. This is consistent with the Stigler hypothesis in that young industries are growing faster and are thus more integrated than mature industries. Tucker and Wilder (1977) examined 54 manufacturing firms using an adjusted value-added to sales ratio and found a U-shaped relationship of vertical integration to the age of the firm. This is consistent with the Stigler hypothesis if the reason why firms are young (old) is that they are in industries which are young (old) and thus have low demand. On the other hand, Stuckey (1983) argues that the increased forward integration by aluminum refiners into fabrications during 1955–78 is inconsistent with the Stigler hypothesis in light of the general growth of the aluminum market.

7.4. Models with demand fluctuations

Perry (1984) builds upon a point made by Oi (1961) to construct a very simple vertical equilibrium. Assume that price fluctuations occur in an intermediate good market because of random exogenous net demand. Oi pointed out that when competitive buyers or sellers can respond ex post to price fluctuations, their expected profits exceed the profits they would earn at the expected price. Buyers cut their losses by reducing purchases of the input and reducing production of the final good when the input price is high; whereas, they profit by increasing input purchases and production when the input price is low. Sellers behave in the opposite fashion. Thus, both buyers and sellers can increase their profits by riding the highs and lows of the market. Even if firms were averse to fluctuating returns, that aversion would have to be very strong in order to overcome the increase in expected profits from price fluctuations. To generate the price fluctuations, Perry makes the simplifying assumption that there is an exogenous random net demand for the input. To further simplify the model, the net demand is assumed to be independent of the input price with an expected value of zero.

From the Oi insight, a buyer and seller would reduce their combined profits if they mutually agreed to exchange a fixed quantity of the input and not otherwise participate in the input market. Thus, there must be economies from the synchronization of production in order for any buyer and seller to vertically merge. These economies might be production economies from coordinating the two stages and eliminating the identity of the intermediate input. They might also arise from reductions in transaction costs as a result of not participating in the input market. Perry simply assumes that these economies augment the profits of integrated firms in the form of a fixed subsidy.

This structure can produce a vertical equilibrium in which some buyers and sellers merge together as integrated firms, while the other buyers and sellers

remain independent. Assume the gains from integration are large enough to induce the first buyer and seller to merge. In the process, they respectively withdraw their stable demand and supply from the open market in the input. As a result, the random exogenous net demand becomes a relatively larger factor in the input market. This amplifies the price fluctuations and increases the expected profits of the independent buyers and sellers. Additional buyers and sellers merge until the gains from synchronization no longer dominate the expected increment in profit from riding the highs and lows of the input market. Thus, even though all buyers and sellers are identical, some will choose to vertically merge while others will remain independent. This finding is not easily reversed by generalizations such as risk aversion and costly storage.

This vertical equilibrium suggests a partial explanation of the more general observation that many industries have a fringe of specialized firms which operate at only one stage. Such firms often adopt more flexible production technologies in order to be more responsive to volatile intermediate good markets. For example, most petroleum products move through integrated channels, but the industry also contains very successful independent crude producers and gasoline retailers. Szenberg (1971) describe a similar story in his case study of the Israeli diamond industry. Although the cutting and polishing of diamonds is primarily handled by integrated firms, subcontractors are employed in the various stages partially to absorb fluctuations in demand.

Green (1974) posits a model of vertical equilibrium with an interesting contrast to the Perry model. A fluctuating exogenous demand and an inflexible price in the intermediate market causes rationing for either upstream or downstream firms. As a result, these firms would like to integrate with one another. However, Green assumes diseconomies of integration. These could arise from either diseconomies of vertical scope or diseconomies of firm size. If the rationing is severe enough, the first upstream and downstream firms will integrate despite these diseconomies. However, this amplifies the severity of the rationing for the other independent firms. This process eventually causes the other upstream and downstream firms to integrate. Thus, the vertical equilibrium is either no integration by any firms or complete integration by all firms. No internal vertical equilibrium exists.

Although there is no explanation for the fixed intermediate price, rationing makes the intermediate market unstable in an undesirable way. This is consistent with the informal stories about vertical integration being a method of avoiding the risks of using the market. Green needs diseconomies of integration so that vertical integration does not automatically arise. The intermediate market in Perry's model is also unstable. But in contrast with Green's model, this is a benefit for disintegrated firms because they can adjust their production. Moreover, Perry needs economies of integration in order to generate the vertical equilibrium.

8. Measurement and trends

8.1. Input measures

We can recognize vertical integration in the context of a specific firm. However, if vertical integration could be measured, then we could also compare firms in the same industry, compare industries with one another, or examine firms and industries over time. The literature on the measurement of vertical integration has developed separately from the theoretical literature previously discussed. The reason for this is that the theory has typically focused upon forward or backward integration by a single firm, while the measurement literature has typically focused upon broader questions concerning the trends of vertical integration.

In Section 1, vertical integration was characterized in terms of the *quantities* transferred internally by a firm operating in two stages of production. Quantity measures are feasible in some industries, the best example being the self-sufficiency ratio used to measure backward integration into crude production by petroleum refiners. However, it is frequently difficult to obtain information on the internal and market exchanges of firms. Thus, other measures have been developed using information that is publicly available. In particular, input measures based upon either capital or labor have been suggested.

Chapman and Ashton (1914) examined the textile industry in both England and other countries. They obtained data from published directories on textile firms. The manufacture of cloth was composed of two distinct stages of production: spinning and weaving. Spindles were employed in the upstream stage for spinning, and looms were employed in the downstream stage for weaving. The data consisted of the number of spindles and looms in each firm. Chapman and Ashton were then able to examine the size distributions of firms which do only spinning, do only weaving, and do both. Thus, physical capital is the measure of vertical integration.

The data indicate that there was a "typical ratio of looms to spindles". This would suggest a synchronization of the output of spindles to the input of looms. However, it is not clear that firms which combined spinning and weaving were integrated in the sense of substituting internal exchange for market exchange. The data represent the equipment stocks of firms, and not the "counts" of yarn transferred from the spinning stage to the weaving stage. Moreover, Chapman and Ashton report that "not infrequently" a combined firm would sell the output of yarn from its spinning and buy the input of yarn for its weaving. Thus, some textile firms may have been simply vertical combinations. This illustrates the potential problem of measuring vertical integration by using the extent to which firms participate in the various stages of production and distribution.

Gort (1962) measures vertical integration in terms of employment rather than equipment. Gort defines the "major" activity of a firm as the one with the most employment. "Auxiliary" activities are the other activities in neighboring stages of production or distribution. The measure of integration is the ratio of employment in auxiliary activities to total employment. From his sample of 111 manufacturing firms in 1954, he finds differences in the degree of integration across 13 manufacturing sectors. The petroleum industry stands out with two-thirds of its employment in auxiliary activities outside of refining, while all of the other sectors were below one-third. This measure has the same potential problem as an asset measure in that it does not necessarily reflect internal transfers.

The measure of quantities transferred internally is certainly the appropriate measure if we are considering vertical integration which arises from technological or transactional economies. However, it may not be an ideal measure in models based upon market imperfections. The reason is that vertical integration in these models leads to different production decisions, thereby confounding the definition with the effects. As a result, input measures can be useful in some theoretical models, as long as it is also assumed that the intermediate input is transferred internally.[43]

8.2. Measures and trends in the twentieth century

The historical discussions of the vertical integration and the growth of modern industry must often rely upon an informal assessment of the extent to which firms participate in various stages. Thorp (1924) and Crowder (1941) classified firms in major industrial groups by their "functions" or types of activities. Several of the classifications involve the combination of vertically related activities within firms. Although this work provided only a picture of U.S. industry at particular points in time, it was the predecessor to later classification schemes designed to examine trends in vertical integration.

Livesay and Porter (1969) examined over 100 major manufacturing firms and classified them according to whether they were integrated backward into the extraction of raw materials or integrated forward into wholesaling or retailing. This procedure has the problem of deciding when partial integration becomes

[43] For example, in Perry (1978c), the monopsonist who partially integrates backward by acquiring suppliers would expand production of the input from these subsidiaries and contract purchases from independent suppliers. Thus, the percentage of input usage from the subsidiaries exceeds the percentage of ownership over the suppliers of the input. It would then seem that ownership of the sources of supply is the relevant measure of integration, and that the percentage of input usage from internal sources is simply a consequence of that vertical integration. This issue is not apparent in most models because only complete integration is examined.

integration for the purpose of classification. Moreover, if the extent to which the industry is integrated is measured by the percentage of its firms which are operating in the other stages, then the impact of decisions by small firms is weighed too heavily relative to those by large firms. Livesay and Porter claim that "variations in integration levels which do exist within industry groups are not, in general, proportionate to the size of firms in the group". As such, their classifications for 6 years in the period 1899–1948 do suggest a trend toward forward integration into wholesaling for a number of manufacturing groups and a trend toward forward integration into retailing for the automobile related groups (petroleum, rubber, and transportation equipment).

Adelman (1955) proposed measuring vertical integration by the ratio of value-added to sales. Value-added equals sales less the costs of materials and intermediate inputs, or equivalently the payments to labor and capital. This measure would partially reflect the difference between integration and combination. The value-added would be the same in both cases, but the sales revenue would be larger under vertical combination because of the sales of the intermediate input. Unfortunately, vertical combination does increase the ratio as if some vertical integration had occurred.[44] But if vertical combination is not prevalent, the ratio of value-added to sales can be useful in comparing the extent of integration between similar firms in a given industry at a particular point in time (also using the same technology and facing the same factor prices). This is the primary emphasis of Adelman's research.

When used for measuring aggregate trends in vertical integration, the value-added to sales ratio has other theoretical problems.[45] However, Adelman presents some illustrative calculations for the entire manufacturing and corporate sector of the economy. He finds that the ratio exhibits no trend for vertical integration. Laffer (1969) confirms this finding for broad subsectors of manufacturing over the period 1929–65. In contrast, Tucker and Wilder (1977) find a small but statistically significant trend toward integration in the weighted ratios of 377 four-digit SIC manufacturing industries over the period 1954–72.

Maddigan (1981) objects to the characterization of vertical integration in terms of sequential stages of production. Instead, she proposes a measure derived from a Leontief input–output matrix for the U.S. economy. Each firm is classified into the various industries in which it participates. The measure of integration is then designed to increase when the firm either (1) participates in more industries

[44] Consider Adelman's example (p. 282) in which the primary producer and the manufacturer each contribute the same value-added. The manufacturer alone would have a ratio of 0.5, while the integrated manufacturer would have a ratio of 1.0. However, a simple combination of the primary producer and manufacturer would have a ratio of 0.67.

[45] Profitability is an important component of value-added, but it varies considerably over the business cycle. See Barnes (1955). To remedy this problem, Tucker and Wilder (1977) suggest an adjusted index which deletes net income and income taxes from the numerator and denominator of the ratio.

having input or output flows between each other or (2) participates in industries having greater input or output flows between themselves. Although creative, this approach cannot capture differences in integration between firms unless they operate in different sets of industries. Moreover, this approach does not allow an independent measure of integration for an entire industry. Indeed, Maddigan must use the average of the measures over the sample of firms in order to conclude that there has been some trend toward integration over the period 1947–72.

If the data were available on the quantities of internal versus external exchanges, one could construct measures of vertical integration similar to the of measures of horizontal integration [see Dirrheimer (1980)]. A vertical Herfindahl–Hirschman Index has recently been introduced into the *Vertical Restraint Guidelines* (1985) issued by the Antitrust Division of the Department of Justice. This index is simply a Herfindahl–Hirschman Index on the firms in an industry which are employing a vertical control. Some problems would arise in applying this measure to vertical integration. In particular, it may be difficult to handle partial integration or to deal with industries having multiple inputs into which firms can integrate.

8.3. Trends prior to the twentieth century

Porter and Livesay (1971) argue that vertical integration in the United States prior to the Civil War was generally backward integration by merchants into the new manufacturing industries. This thesis is based upon the observations that wealth during this period resided with the merchants and that capital markets were non-existent or very imperfect. Thus, backward integration was the source of financing for industrial development. Before examining the evidence produced by Porter and Livesay, we briefly point out that their thesis may have also been accurate in European history.

International trade virtually ceased after the fall of the Roman Empire. Saracen pirates controlled the Mediterranean while Norse raiders dominated the Atlantic coast. But beginning in the eleventh century, merchants arose in the coastal cities to resume international trade.[46] The first merchant cities were Venice on the Mediterranean and Bruges in Flanders on the Atlantic. The merchants amassed wealth from the profits on the local products exported to other lands. Thus, the development of a local extractive or manufacturing industry became an important element of the success of local merchants. For example, Venetian merchants exported wheat, wine, wood, and salt from the surrounding areas of Italy.[47] Merchants often financed the local industries in a

[46]See Pirenne (1974, p. 90).
[47]See Pirenne (1974, pp. 85–87).

form of early banking. But merchants also became more directly involved in local industries by acquiring machinery, hiring local craftsmen, and managing production.

The textile industry is replete with examples of merchants integrating into the production of export products. Local craftsmen from rural areas were attracted to the cities to produce export products. For example, the merchants of Bruges imported wool from England and produced fine woven dyed cloths for export.[48] The Venetian merchants also imported wool and then financed and supervised the production of cloth through a putting-out system. This pattern apparently continued into the sixteenth century in many parts of Western Europe. Flemish merchants continued to produce worsted yarn spun from long-staple wool, and the merchants of Lyon produced silk fabric from imported raw silk.[49]

We now turn to the historical evidence of Porter and Livesay (1971). During the early 1800s, merchants in the United States specialized in the export of specific commodities. Most notably, certain merchants specialized in the export of cotton. Although these merchants did not directly produce cotton, they usually advanced credit to the planters. Other merchants specialized in the distribution of imported textiles. Some of these merchants then financed the New England textile industry. Typically, the merchants funded the construction of the mills and engaged a British manufacturing expert to manage the operations. The merchants would then set up a selling house to distribute the textiles produced by their mills. The selling house provided working capital and took the risk of unsold production.[50]

In the iron industry, specialized merchants served as the intermediaries between the producers of pig iron located near the ore deposits and either the foundries located near cities of the rolling mills located near water power. However, prior to the Civil War, the pig iron furnaces were integrated with the rolling mills into new companies financed by the iron merchants of Pennsylvania.[51] The new companies supplied rail for the rapidly growing railroad industry, and became the forerunners of the current steel corporations.

[48]See Pirenne (1974, pp. 153–155).

[49]Silver (1984, pp. 98–103) cites several other examples.

[50]Porter and Livesay (1971) cite the example of Francis Cabot Lowell who founded the Boston Manufacturing Company in 1813 along with other Boston textile merchants. The company constructed and operated mills on the Merrimack River. Later in 1828, B.C. Ward and Company was formed as a selling house for the entire output of the mills. See also Temin (1987) for a detailed examination (with an extensive bibliography) of vertical integration in the early textile industry of both Great Britain and the United States. Temin takes issue with some transaction cost explanations of this history, and suggests that one alternative explanation for vertical integration was the existence of capital in the more mature neighboring stages of production.

[51]Coal was substituted for charcoal as fuel for the furnaces, and steam power was substituted for water power.

These and other examples lead Porter and Livesay to conclude that the vertical integration of distribution into manufacturing in early American business was a consequence of the primitive state of capital markets. Capital resided with merchants prior to the Civil War, but the investment opportunities were in the manufacturing of new products. Financial institutions such as banks and insurance companies were generally unwilling to make loans without collateral such as government bonds or real estate. With the exception of the railroads, manufacturers could not generally raise money by public offerings of securities. Thus, merchants became direct participants in the new enterprises, merging the source of investment funds with the capital collateral.

During the Civil War, manufacturers prospered and were able to end their dependence upon merchant distributors for either investment or working capital. Indeed, the profits from manufacturing in the late nineteenth century were an important source of investment funds. Porter and Livesay discuss many cases during this period in which manufacturers integrated either backward into extraction or forward into distribution (see Subsection 5.3). But the question remains to what extent this integration was determined by the need for financing. On the one hand, the nineteenth century witnessed the gradual development of capital markets and contract law.[52] These developments should have reduced the financial motive for vertical integration, in that firms could contract for independent financing. On the other hand, the case studies indicate that vertical integration did not cease and was typically initiated by the stage with the largest and most profitable firms, manufacturers or large retailers. Moreover, vertical quasi-integration developed as a form of retail financing. Thus, the importance of retained earnings as a determinant of modern vertical integration remains an intriguing question.

9. Public policy

Vertical expansion has not been generally found to be a violation of the U.S. antitrust laws. On the other hand, vertical merger has been viewed much less favorably, at least until recently. Although neither has been held to be an unreasonable restraint of trade in violation of §1 of the Sherman Act, both have been held to be exclusionary practices in monopolization cases under §2. But the primary difference in treatment arises from §7 of the Clayton Act which condemns vertical mergers that substantially lessen competition.

[52] Horwitz (1977, ch. VI) examines the symbiotic development of capital markets and contract law during the nineteenth century.

9.1. Section 2 of the Sherman Act

Vertical expansion by dominant firms has been condemned in a few cases as an exclusionary practice in the acquisition or maintenance of monopoly at the primary stage of production. Alcoa's pre-emptive backward integration into bauxite and electric power generation was found to be an exclusionary practice in that it created barriers to entry protecting Alcoa's monopoly at the refining stage.[53] However, no vertical divestiture was ordered. A & P was also held to have violated §2 for exclusionary practices related to its backward integration into wholesale food production and distribution.[54] A & P was said to have abused its buying power by threats to boycott suppliers or to integrate into processing. In addition, its upstream buying subsidiary discriminated against independent grocers in favor of A & P stores in supplying produce.[55] Other cases are similar to A & P in that vertical integration occurred in conjunction with exclusionary practices such as refusals to deal or price discrimination. Thus, despite *Alcoa*, vertical expansion is probably not a sufficient exclusionary practice by itself for antitrust liability under §2.

Vertical divestiture was the horizontal remedy in two major monopolization cases under §2. The Standard Oil Trust was ordered to divest in 1911.[56] The corporate members severed from the Trust were generally specialized in crude production, pipeline transportation, refining, or marketing. More recently, the Bell System has been vertically divested as a consequence of a 1982 consent decree settling the antitrust suit initiated by the Justice Department in 1972.[57] AT & T divested the Bell operating companies supplying local service but continues to supply long-distance service and produce equipment.[58]

Several cases under §2 deal with forward integration as an extension of upstream monopoly into downstream stages of distribution. The upstream monopoly may be legal or unchallenged so that the §2 violation is monopolization of the downstream stage. These cases present less deference toward vertical integration by a monopolist, but one important reason is that they involve vertical acquisitions as well as vertical expansion. An early example is Kodak's forward integration into the distribution of its photographic supplies by acquiring whole-

[53] *U.S. v. Aluminum Company of America* (2nd Circuit, 1945, 148 F.2d 416).

[54] *U.S. v. New York Great Atlantic and Pacific Tea Co.* (7th Circuit, 1949, 173 F.2d 79).

[55] Adelman (1949b) thoroughly criticizes the economic analysis employed by the Justice Department and adopted by the court. The case has also been severely criticized by legal scholars. Although A & P was the largest grocery chain, its national market share was only 10 percent. The Justice Department eventually abandoned its attempt to split A & P into seven regional grocery chains.

[56] *Standard Oil Co. of New Jersey v. U.S.* (1911, 221 U.S. 1).

[57] *U.S. v. AT & T Co.* (D.D.C., 1982, 552 F.Supp. 131).

[58] The offspring of Standard Oil partially reintegrated into the other stages [see Johnson (1976)]. Similarly, the Bell operating companies are seeking court permission to re-enter the long distance and equipment markets, and AT & T is attempting to re-enter certain types of local service.

sale houses in 1910. Kodak's tactic of discontinuing the usual wholesale discounts to an independent distributor who had refused to be acquired was held to be an illegal attempt to monopolize the distribution stage.[59]

Forward integration into exhibition by the major producers of motion pictures was challenged by the Justice Department in 1938 as a conspiracy to monopolize the distribution of feature films. The Supreme Court stated that vertical integration can violate §2 of the Sherman Act "if it was a calculated scheme to gain control over an appreciable segment of the market and to restrain or suppress competition, rather than an expansion to meet legitimate needs".[60] Other exclusionary practices against independent theaters were also examined in the case, but the key remedy ordered in 1949 was vertical divestiture.[61]

More recently, Otter Tail Power, an electric power company, was held to have violated §2 by refusing to sell power at wholesale or to transmit power from other sources to retail municipal electric systems within its territory.[62] These exclusionary practices forced over 90 percent of the municipalities to grant Otter Tail the franchise to perform retail distribution within their boundaries. Vertical divestiture was not essential because by enjoining the exclusionary practices, a municipality could acquire Otter Tail's retail distribution system when the franchise expired.

In each of these cases, a price or supply squeeze was used to either acquire or maintain a dominant position in the downstream stage. The price squeeze is a necessary consequence of forward integration to effectuate price discrimination (see Subsection 2.2). But in these Sherman Act cases, the price squeeze is viewed as a device to extend an existing upstream monopoly into a downstream stage. Bork (1978) argues that there is no additional welfare loss from the mere extension of monopoly or oligopoly into neighboring stages. Indeed, vertical integration can achieve some efficiencies. This view was reflected in a recent Court of Appeals decision which rejected a §2 challenge to forward integration by a monopoly newspaper. The Kansas City Star discontinued distributing its newspaper through independent contract carriers, and began delivering the newspaper itself. The court recognized that forward integration would eliminate the successive monopoly problem inherent in the territorial arrangements with independent carriers.[63]

The requirement of substantial market share prevents §2 from being an effective method of challenging vertical integration by firms in less concentrated

[59] *Eastman Kodak Co. of New York v. Southern Photo Materials Co.* (1927, 273 U.S. 359).

[60] *U.S. v. Paramount Pictures* (1948, 334 U.S. 131).

[61] Block booking and circuit renting were also enjoined in order to force picture-by-picture, theater-by-theater rentals. *U.S. v. Paramount Pictures* (S.D.N.Y., 1949, 85 F.Supp. 881).

[62] *Otter Tail Power Co. v. U.S.* (1973, 410 U.S. 366).

[63] *Paschall v. The Kansas City Star Co.* (8th Circuit, 1984). Note that maximum resale price maintenance, an alternative solution to the successive monopoly problem, was held to be a violation of §1 in a similar newspaper case. See *Albrecht v. Herald Co.* (1968, 390 U.S. 145).

markets. In 1946, U.S. Steel, through its West Coast subsidiary Columbia Steel, acquired Consolidated Steel, a West Coast steel fabricator with 11 percent of the regional market. The merger would have given U.S. Steel 24 percent of the regional market in fabricated steel products. The Justice Department argued that the acquisition would foreclose Consolidated's demand for rolled steel from the competitors of U.S. Steel. The Supreme Court responded that "vertical integration, as such without more, cannot be held violative of the Sherman Act".[64] The market foreclosure must "unreasonably restrict the opportunities of competitors to market their product", probably meaning it must rise to the level of monopolization.

9.2. Section 7 of the Clayton Act

Prior to 1950, §7 of the Clayton Act did not apply to vertical vergers. But the concern over mergers after World War II and the *Columbia Steel* limitations on the Sherman Act lead to the Celler–Kefauver amendments to §7 in that year. Vertical mergers became illegal "where in any line of commerce..., the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly". Both of these criteria are weaker than the judicial requirements under §2 of the Sherman Act. Thus, since 1950, vertical mergers have been challenged under the Clayton Act rather than the Sherman Act.

The predominant legal theory employed to challenge vertical mergers is market foreclosure.[65] Under the market foreclosure theory, vertical merger harms competition in both stages by denying competitors access either to one of their suppliers or to one of their buyers. Market foreclosure was employed in the three major vertical merger cases to reach the Supreme Court after 1950.

The Court condemned the 23 percent stock interest that duPont had held in General Motors since 1919.[66] During the 1950s, General Motors manufactured about 50 percent of the automobiles in the domestic market, while duPont supplied as much as two-thirds of the finishes and half of the fabrics used by General Motors. Thus, the actual foreclosure was significant. This was not true for the acquisition of Kinney by Brown Shoe in 1955. Brown Shoe accounted for only 5 percent of the domestic shoe production, while Kinney's chain of family shoe stores accounted for only 1 percent of the domestic shoe sales. A trend toward vertical integration and foreclosure were employed by the Court to condemn this merger.[67] Finally, foreclosure was a consideration in condemning

[64]*U.S. v. Columbia Steel Co.* (1948, 334 U.S. 495).

[65]Adelman (1949a) has a nice discussion of the early antitrust fallacies concerning vertical integration.

[66]*U.S. v. E.I. duPont de Nemours & Co.* (1957, 353 U.S. 586).

[67]*Brown Shoe Co. v. U.S.* (1962, 370 U.S. 294).

Ford's acquisition of Electric Autolite, a spark plug manufacturer, in 1961.[68] Ford had 25 percent of the domestic automobile market, while Electric Autolite manufactured 15 percent of the domestic spark plugs.

In the late 1960s and early 1970s, the Federal Trade Commission employed §7 to block or dissolve a number of vertical acquisitions of ready-mix concrete firms by cement manufacturers. In 1966, an FTC staff report argued that these vertical acquisitions foreclosed markets and caused a "chain reaction" of defensive vertical mergers. The effect of these vertical mergers was to increase the barriers to entry and concentration in the cement industry.[69] The numerous comments on the FTC report focused upon the excess capacity in the cement industry as the key factor explaining the subsequent forward integration.[70] This excess capacity arose not from the construction of more plants having the same size and efficiency, but rather from the construction of larger plants which reduced the average cost per barrel by at least 50 percent [see Peck and McGowen (1967)]. Between 1950 and 1964, the demand for cement much less than doubled while the minimum efficient scale of a cement plant at least tripled. This suggests a rationalization of the cement industry in which vertical acquisition was employed by new efficient firms to hasten the inevitable trend and by old inefficient firms in a futile attempt to prevent it. If so, the vertical acquisitions were a substitute for price competition in reducing the market share of smaller inefficient plants, and should have had no long-run efficiency or strategic significance.[71] Similarly, McBride (1983) has argued that the vertical acquisitions and foreclosure were non-price competition strategically designed to increase capacity utilization and market share in the short run while postponing price competition among the cement oligopolists.[72]

The original Department of Justice *Merger Guidelines* (1968) incorporated the foreclosure theory, but not to the extent of the *Brown Shoe* case. The *Guidelines* suggested that vertical mergers between an upstream firm with more than 10 percent of its market and a downstream firm with more than 6 percent of its market would be challenged by the government. However, the foreclosure theory has been strongly criticized by both economists and lawyers.[73] The basic point is that vertical foreclosure is a definition rather than a theory. Since vertical

[68] *Ford Motor Co. v. U.S.* (1972, 405 U.S. 562).

[69] See Mueller (1969). Whalen (1969) questioned this strategic theory of vertical foreclosure because of the ease of entry into the production of ready-mix concrete. Allen (1971) questioned the profitability of vertical foreclosure by the cement manufacturers.

[70] Peck and McGowen (1967) argued that forward integration by cement manufacturers was a natural investment opportunity given the excess capacity and a falling rate of return on cement during the early 1960s. Wilk (1968) pointed out that excess capacity would only explain why cement manufacturers did not reinvest in their own stage of production.

[71] Inefficiencies could arise in the short-run if the cement manufacturers with suboptimal size plants were able to survive longer as a result of vertical acquisitions.

[72] McBride provides some empirical evidence of short-run rigidity in the price of cement.

[73] See Comanor (1967), Peltzman (1969), and Bork (1978).

integration is the substitution of internal exchange for market or contractual exchange, foreclosure would naturally occur if either partner dealt with other firms prior to the merger. This and other criticisms have made an impact on recent decisions in some Circuit Courts of Appeal. For example, the Second Circuit refused to enforce an FTC divestiture order of Fruehauf, the largest manufacturer of truck trailers with 25 percent of the market, from Kelsey-Hayes which had 15 percent of the market for truck and trailer wheels.[74] The court rejected the foreclosure theory absent very high market concentration in one of the two stages.

Criticism of the foreclosure theory has also resulted in a change in enforcement policy by the Department of Justice. Under Assistant Attorney General William Baxter, the Antitrust Division of the Department of Justice revised its *Merger Guidelines* in 1982 and abandoned the market foreclosure theory. Bork (1978), Posner (1976), Baxter, and others have argued that vertical mergers often generate important efficiencies and can present no competitive problem unless one stage is highly concentrated. The current *Merger Guidelines* (1984) now reflect this approach. The Department is unlikely to challenge the vertical merger by a firm unless its stage of the industry has a Herfindahl–Hirschman Index above 1800. Even then, the Department may decline to challenge a vertical merger when there is evidence of substantial economies such as better coordination of production and distribution.

The *Guidelines* also fail to recognize a merger "trend" as being relevant for challenging any given vertical merger in an industry. The legislative history and language of §7 demonstrate some concern about industry trends, i.e. the issue of "incipient" monopoly. Moreover, a perceived trend of vertical mergers was an important factor in condemning the merger in the *Brown Shoe* case and the mergers in the cement industry. Although it may be difficult to incorporate such considerations into the *Guidelines*, it does seem that there is a valid economic concern. The first firms to integrate into neighboring stages reduce the number of alternative sources for other firms at either stage. This "thinning" of the market can increase the costs of market or contractual exchange (see Subsection 2.3). Subsequent integration by other firms then becomes more likely. This raises two issues. First, it may be inaccurate to formulate public policy from economic theories that focus only upon the production, exchange, or market problems of a particular buyer and seller. Second, the private incentives for vertical merger may not generate a vertical industry structure which is socially optimal. This is another reason why research on vertical equilibrium should be enlightening (see Section 7).

[74]The actual foreclosure of this merger was less than 6 percent. *Fruehauf Corp. v. FTC* (2nd Circuit, 1979, 603 F.2d 345).

The *Guidelines* have clearly reduced the likelihood of a Justice Department challenge to a vertical merger. For example, the 1981 vertical merger of the chemical company duPont and the petroleum company Conoco would not have been challenged under the *Guidelines*. Unlike the 1950s and 1960s, current vertical merger enforcement is much closer to the deference accorded vertical expansion. However, the standards articulated in the *Guidlines* have not been incorporated into court decisions to narrow the existing precedents under §7.[75] The *Ford Motor* case remains the relevant precedent. In that case, the Supreme Court condemned Ford's acquisition of Electric Autolite on the basis that it would increase barriers to entry in the spark plug industry and eliminate Ford as a major potential entrant. Even though the automobile market was concentrated, it seems unlikely that this merger would have triggered the criteria for challenge under the *Guidelines*.

The *Guidelines* do recognize three major competitive problems of vertical mergers in concentrated industries. First, forward mergers into retailing may facilitate collusion at the manufacturing stage by making it easier to monitor prices or by eliminating a "disruptive buyer". Second, vertical mergers may enhance barriers to entry into the primary industry if entrants must operate at both stages in order to be competitive with existing firms and if entry at both stages is substantially more difficult than entry at one stage. Finally, vertical mergers may have an adverse competitive impact by eliminating specific potential entrants who could integrate by vertical expansion rather than merger. Each of these avenues of challenge are further circumscribed by market structure conditions which narrow their force.

9.3. Public policy in the United Kingdom and the European Economic Community

Similar to U.S. antitrust law, European "competition" law makes a distinction between vertical expansion and vertical merger. We illustrate the parallels with a brief discussion of the United Kingdom (U.K.) and the European Economic Community (EEC). The discussion relies heavily upon Merkin and Williams (1984).[76]

In the United Kingdom, complaints about violations of the competition laws are investigated by the Office of Fair Trading (OFT). The Director General of the

[75] Private parties can challenge vertical mergers but they must also prove a private injury (for damages) or a threat of private injury (for an injunction). Moreover, the private injury must be of the type which the antitrust laws were designed to prevent and which flows from the anticompetitive consequences of the merger. These requirements limit the ability of private parties to challenge mergers which would have anticompetitive consequences for consumers. See *Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc.* (1977, 429 U.S. 477) and *Cargill, Inc. v. Monfort of Colorado, Inc.* (1986, 479 U.S. 104).

[76] See also Korah (1982).

OFT may then refer a case to the Monopolies and Mergers Commission (MMC) which has the responsibility for evaluating monopolies, mergers, and most vertical practices.[77] The MMC employs a "public interest" standard to promote price and non-price competition. The non-price considerations include product quality and variety, the development of new techniques and products, and entry. The MMC will report to the Secretary of State (Department of Trade and Industry) on whether the laws were violated and will make recommendations concerning the remedies which should be sought.

In the European Economic Community, competition law arises from Articles 85 and 86 of the Treaty of Rome creating the EEC in 1957. Vertical integration is covered by Article 86 which prohibits "abuse of a dominant position".[78] The European Commission is the executive body of the EEC, and the administration of competition policy is implemented by one of its departments (Directorates-General IV). The DG IV investigates possible violations, and the Commission can issue cease and desist orders or impose fines. The firms subject to these penalties can then appeal to the European Court of Justice.

In the U.K., vertical expansion by a dominant firm has been challenged when it results from a vertical price or supply squeeze. The MMC carefully investigated the complaints of independent bakers concerning a price squeeze by the dominant millers of flour. The millers were forward integrated into baking, but the MMC found no proof of a squeeze designed to increase the market share of their baking subsidiaries.[79] In a similar case, the MMC investigated British Oxygen, the dominant manufacturer of machinery used to produce oxygen. British Oxygen was forward integrated into the production of oxygen, and the MMC criticized the company for limiting the supply of machinery available to independent producers of oxygen.[80]

In the EEC, vertical expansion has been condemned if it results from a refusal to deal. Commercial Solvents Corporation (CSC) had a dominant position in the world production of aminobutanol, an input in the production of certain drugs. Zoja manufactured ethambutol from aminobutanol, and had purchased the latter from CSC until switching suppliers in 1970. CSC unsuccessfully attempted to takeover Zoja, and thereafter began producing ethambutol itself. When Zoja lost its alternative sources of aminobutanol, CSC refused to sell the input to Zoja. The European Court of Justice held that this was an abusive practice in violation

[77]The Fair Trading Act of 1973 vests jurisdiction over monopolies and mergers, while the Competition Act of 1980 vests jurisdiction over most vertical practices. Horizontal agreements and resale price maintenance are referred to the Restrictive Practices Court.

[78]Article 85 is concerned with horizontal agreements and practices.

[79]*Flour and Bread (1976–77)* H.C. 412.

[80]Procedural reasons prevented the MMC from recommending remedial measures. *Industrial and Medical Gases (1956–57)* H.C. 13.

of Article 86.[81] In a related case, a manufacturer of cash registers refused to supply spare parts for its products to an independent service firm. The Advocate General, an assistant to the Court, agreed with the Commission that this practice was abusive. However, the Court expressed no opinion on this substantive issue after finding that interstate trade was not affected, as required by Article 86.[82] These cases suggest that vertical expansion is clearly not sufficient to violate the competition laws unless it also involves practices designed to disadvantage or exclude competitors. This is similar to the prevailing view of vertical expansion in the United States under §2 of the Sherman Act. Vertical integration is not illegal itself, but rather it provides an opportunity and maybe an incentive for other practices deemed undesirable.

European competition law with respect to vertical mergers is less developed and apparently less controversial than in the United States. Indeed, in the EEC, it is unresolved whether vertical mergers are subject to Article 86.[83] In the U.K., OFT investigations of vertical mergers typically examine the issue of foreclosure. However, the MMC has recognized the weaknesses of the foreclosure theory and has not employed it to condemn vertical mergers. Berisford, the dominant distributor of sugar in the U.K., proposed to acquire British Sugar, the leading U.K. producer of sugar. Tate and Lyle, the only other producer of sugar, objected to the possible foreclosure of its distribution channel. The MMC approved the merger, but subject to the condition that Berisford continue to distribute Tate and Lyle products.[84]

The MMC has rejected a vertical merger on the grounds that it would eliminate a potential entrant into R&D. Boots, a pharmaceutical company with an R&D laboratory, attempted to acquire Glaxo Group, a pharmaceutical manufacturer with a major R&D laboratory. This was viewed as a backward merger, and was disallowed by the MMC on the grounds that Boots might develop its own R&D capability.[85] Since this case also reflects a strong interest in encouraging R&D in the U.K., it is not clear that it reflects a strong precedent for challenging other types of vertical mergers. However, the focus upon potential competition, rather than foreclosure, is similar to the modern analysis of vertical mergers in the United States.

[81] *Instituto Chemioterapico Italiano and Commercial Solvents v. Commission.* Cases 6 and 7/73 (1974) E.C.R. 707.

[82] *Hugin Kassaregister v. Commission.* Case 22/78 (1979) E.C.R. 1869. In the U.K., an identical practice was prevented by an OFT investigation. A manufacturer of catering equipment had refused to supply independent service firms with spare parts for its equipment. See *Still* (OFT, July 1982).

[83] The European Court of Justice has held that horizontal mergers are subject to Article 86. See *Europembellage and Continental Can Co. v. Commission.* Case 6/72 (1972) E.C.R. 157.

[84] *Berisford Ltd. and British Sugar Corporation Ltd. (1980–81)* H.C. 241.

[85] *The Boots Company Ltd. and Glaxo Group Ltd. (1971–72)* H.C. 341.

References

Adelman, M.A. (1949a) 'Integration and antitrust policy', *Harvard Law Review*, 63:27–77.

Adelman, M.A. (1949b) 'The A & P case: A study in applied economic theory', *Quarterly Journal of Economics*, 63:238–257.

Adelman, M.A. (1955) 'Concept and statistical measurement of vertical integration', in: G.J. Stigler, ed., *Business concentration and price policy*. Princeton, N.J.: Princeton University Press.

Alchian, A.A. and Demsetz, H. (1972) 'Production, information costs, and economic organization', *American Economic Review*, 62:777–795.

Allen, B.T. (1971) 'Vertical integration and market foreclosure: The case of cement and concrete', *Journal of Law and Economics*, 14:251–274.

Anderson, E. and Schmittlein, D. (1984) 'Integration of the sales force: An empirical examination', *The Rand Journal of Economics*, 15:385–395.

Areeda, P. (1981) *Antitrust analysis*. Boston: Little, Brown and Company.

Armour, H.O. and Teece, D.J. (1980) 'Vertical integration and technological innovations', *Review of Economics and Statistics*, 62:490–494.

Arrow, K.J. (1975) 'Vertical integration and communication', *Bell Journal of Economics*, 6:173–183.

Bain, J.S. (1956) *Barriers to new competition*. Cambridge, Mass.: Harvard University Press.

Barnes, I.R. (1955) 'Concept and statistical measurement of vertical integration: Comment', in: G.J. Stigler, ed., *Business concentration and price policy*. Princeton, N.J.: Princeton University Press.

Baumol, W.J., Panzar, J.C. and Willig, R.D. (1982) *Contestable markets and the theory of industry structure*. New York: Harcourt Brace Jovanovitch.

Bernhardt, I. (1977) 'Vertical integration and demand variability', *Journal of Industrial Economics*, 25:213–229.

Blair, R.D. and Kaserman, D.L. (1978) 'Vertical integration, tying, and antitrust policy', *American Economic Review*, 68:397–402.

Blair, R.D. and Kaserman, D.L. (1983) *Law and economics of vertical integration and control*. New York: Academic.

Blois, K.J. (1972) 'Vertical quasi-integration', *Journal of Industrial Economics*, 20:253–272.

Bork, R. (1978) *The antitrust paradox*. New York: Basic Books.

Burstein, M. (1960) 'The economics of tie-in sales', *Review of Economics and Statistics*, 42:68–73.

Buzzell, R.D. (1983) 'Is vertical integration profitable?', *Harvard Business Review*, 61:92–102.

Calamari, J.D. and Perillo, J.M. (1977) *Contracts*. St. Paul, Minn.: West Publishing.

Carlton, D.W. (1979) 'Vertical integration in competitive markets under uncertainty', *Journal of Industrial Economics*, 27:189–209.

Carlton, D.W. and Perloff, J.M. (1981) 'Price discrimination, vertical integration and divestiture in natural resource markets', *Resources and Energy*, 3:1–11.

Chandler, A.D. (1977) *The visible hand: The managerial revolution in American business*. Cambridge, Mass.: Harvard University Press.

Chapman, S.J. and Ashton, T.S. (1914). 'The sizes of businesses, mainly in the textile industries', *Journal of the Royal Statistical Society*, 77:469–549.

Cheung, S.N.S. (1983) 'The contractual nature of the firm', *Journal of Law and Economics*, 26:1–21.

Cipolla, C.M. (1980) *Before the industrial revolution: European society and economy*, 1000–1700. New York: Norton.

Coase, R. (1937) 'The nature of the firm', *Economica*, 4:386–405.

Comanor, W.S. (1967) 'Vertical mergers, market power, and the antitrust laws', *American Economic Review*, 57:254–265.

Crandall, R. (1968) 'Vertical integration and the market for repair parts in the United States automobile industry', *Journal of Industrial Economics*, 16:212–234.

Crocker, K.J. (1983) 'Vertical integration and the strategic use of private information', *Bell Journal of Economics*, 14:236–248.

Crowder, W.F. (1941) *The integration of manufacturing operations*. Washington, D.C.: T.N.E.C. monograph no. 27.

deChazeau, M.G. and Kahn, A.E. (1959) *Integration and competition in the petroleum industry*. New Haven: Yale University Press.

Department of Justice, Antitrust Division (1984) 'Merger guidelines', in: M. Handler, H.M. Blake, R. Pitofsky and H.J. Goldschmid, eds., *Trade regulation, 1985 supplement*. Mineola, N.Y.: Foundation Press.

Department of Justice, Antitrust Division (1985) 'Vertical restraint guidelines', in: M. Handler, H.M. Blake, R. Pitofsky and H.J. Goldschmid, eds., *Trade regulation, 1985 supplement*. Mineola, N.Y.: Foundation Press.

Dirrheimer, M.J. (1980) 'Vertical integration: Transaction cost advantages versus market power disadvantages', International Institute of Management, Berlin.

Dixit, A. (1983) 'Vertical integration in a monopolistically competitive industry', *International Journal of Industrial Organization*, 1:63–78.

Dixit, A. and Stiglitz, J.E. (1977) 'Monopolistic competition and optimum product diversity', *American Economic Review*, 67:297–308.

Evans, D.S. and Grossman, S.J. (1983) 'Integration', in: D.S. Evans, ed., *Breaking up Bell*. New York: North-Holland.

Federal Trade Commission (1966) *Economic report on mergers and vertical integration in the cement industry*. Washington: U.S. Government Printing Office.

Fellner, W. (1947) 'Prices and wages under bilateral monopoly', *Quarterly Journal of Economics*, 61:503–509.

Flaherty, M.T. (1981) 'Prices versus quantities and vertical financial integration,' *Bell Journal of Economics*, 12:507–525.

Goldberg, V.P. (1976) 'Regulation and administered contracts', *Bell Journal of Economics*, 7:426–448.

Goldberg, V.P. and Erickson, J.R. (1987) 'Quantity and price adjustment in long-term contracts: A case study of petroleum coke', *Journal of Law and Economics*, 30:369–398.

Gort, M. (1962) *Diversification and integration in American industry*. Princeton, N.J.: Princeton University Press.

Gould, J.R. (1977) 'Price discrimination and vertical control: A note', *Journal of Political Economy*, 85:1063–1071.

Green, J.R. (1974) 'Vertical integration and assurance of markets', discussion paper no. 383, Harvard Institute of Economic Research.

Greenhut, M.L. and Ohta, H. (1976) 'Related market conditions and interindustrial mergers', *American Economic Review*, 66:267–277.

Greenhut, M.L. and Ohta, H. (1978) 'Related market conditions and interindustrial mergers: Reply', *American Economic Review*, 68:228–230.

Greenhut, M.L. and Ohta, H. (1979) 'Vertical integration of successive monopolists', *American Economic Review*, 69:137–141.

Grossman, S.J. and Hart, O.D. (1986) 'The costs and benefits of ownership: A theory of vertical and lateral integration', *Journal of Political Economy*, 94:691–719.

Gunther, G. (1980) *Cases and materials on constitutional law*. Mineola, N.Y.: Foundation Press.

Hale, R.D. (1967) 'Cookware: A study in vertical integration', *Journal of Law and Economics*, 10:169–179.

Handler, M., Blake, H.M., Pitofsky, R. and Goldschmid, H.J. (1983) *Trade regulation*. Mineola, N.Y.: Foundation Press.

Haring, J.R. and Kaserman, D.L. (1978) 'Related market conditions and interindustrial mergers: Comment', *American Economic Review*, 68:225–227.

Hart, O. and Holmstrom, B. (1986) 'The theory of contracts', in: T. Bewley, ed., *Advances in Economic Theory*. Cambridge University Press.

Hart, O. and Moore, J. (1988) 'Incomplete contracts and renegotiation', *Econometrica*, 56:755–785.

Hay, G. (1973) 'An economic analysis of vertical integration', *Industrial Organization Review*, 1:188–198.

Helfat, C.E. and Teece, D.J. (1987) 'Vertical integration and risk reduction', *Journal of Law, Economics, and Organization*, 3:47–67.

Horwitz, M.J. (1977) *The transformation of American law, 1780–1860*. Cambridge, Mass.: Harvard University Press.

Hovenkamp, H. (1985) *Economics and federal antitrust law*. St. Paul, Minn.: West Publishing.

Johnson, A.M. (1976) 'Lessons of the Standard Oil divestiture', in: E.J. Mitchell, ed., *Vertical integration in the oil industry*. Washington, D.C.: American Enterprise Institute.
Joskow, P.L. (1985) 'Vertical integration and long-term contracts: The case of coal-burning electric generating plants', *Journal of Law, Economics, and Organization*, 1:33–80.
Kaserman, D.L. (1978) 'Theories of vertical integration: Implications for antitrust policy', *Antitrust Bulletin*, 23:483–510.
Katz, M.L. (1987) 'The welfare effects of third degree price discrimination in intermediate good markets', *American Economic Review*, 77:154–167.
Klein, B., Crawford, R.G. and Alchian, A.A. (1978) 'Vertical integration, appropriable rents, and the competitive contracting process', *Journal of Law and Economics*, 21:297–326.
Korah, V. (1982) *Competition law of Britain and the Common Market*. The Hague: Martinus Nijhoff Publishers.
Laffer, A.B. (1969) 'Vertical integration by corporations: 1929–1965', *Review of Economics and Statistics*, 51:91–93.
Laffont, J.J. (1976) 'More on prices vs. quantities', *Review of Economic Studies*, 43:177–182.
Levin, R.C. (1981) 'Vertical integration and profitability in the oil industry', *Journal of Economic Behavior and Organization*, 2:215–235.
Levy, D.T. (1984) 'Testing Stigler's interpretation of "The division of labor is limited by the extent of the market"', *Journal of Industrial Economics*, 32:377–389.
Levy, D.T. (1985) 'The transactions cost approach to vertical integration: An empirical examination', *Review of Economics and Statistics*, 67:438–445.
Livesay, H.C. and Porter, P. (1969) 'Vertical integration in American manufacturing', *Journal of Economic History*, 29:494–500.
Machlup, F. and Taber, M. (1960) 'Bilateral monopoly, successive monopoly, and vertical integration', *Economica*, 27:101–119.
Maddigan, R.J. (1981) 'The measurement of vertical integration', *Review of Economics and Statistics*, 63:328–335.
Mallela, P. and Nahata, B. (1980) 'Theory of vertical control with variable proportions', *Journal of Political Economy*, 88:1009–1025.
Malmgren, H.B. (1961) 'Information, expectations, and the theory of the firm', *Quarterly Journal of Economics*, 75:339–421.
Mancke, R.B. (1972) 'Iron ore and steel: A case study of the economic causes and consequences of vertical integration', *Journal of Industrial Economics*, 21:220–229.
Mancke, R.B. (1982) 'The petroleum industry', in: W. Adams, ed., *The structure of American industry*. New York: Macmillan.
Masten, S.E. (1984) 'The organization of production: Evidence from the aerospace industry', *Journal of Law and Economics*, 27:403–417.
Mathewson, G.F. and Winter, R.A. (1983) 'Vertical integration by contractual restraints in spatial markets', *Journal of Business*, 56:497–517.
Mathewson, G.F. and Winter, R.A. (1984) 'An economic theory of vertical restraints', *Rand Journal of Economics*, 15:27–38.
Mathewson, G.F. and Winter, R.A. (1986) 'The economics of vertical restraints in distribution', in: J.E. Stiglitz and G.F. Mathewson, eds., *New developments in analysis of market structures*. Cambridge, Mass.: MIT Press.
McBride, M.E. (1983) 'Spatial competition and vertical integration: Cement and concrete revisited', *American Economic Review*, 73:1011–1022.
McGee, J.S. and Bassett, L.R. (1976) 'Vertical integration revisited', *Journal of Law and Economics*, 19:17–38.
McKenzie, L.W. (1951) 'Ideal output and the interdependence of firms', *Economic Journal*, 61:785–803.
McLean, J.G. and Haigh, R.W. (1954) *The growth of integrated oil companies*. Boston, Mass.: Harvard Business School.
McNicol, D.L. (1975) 'The two price system in the copper industry', *Bell Journal of Economics*, 6:50–73.
Merkin, R. and Williams, K. (1984) *Competition law: Antitrust policy in the U.K. and the EEC*. London: Sweet and Maxwell.

Mitchell, E.J. (1976) 'Capital cost savings of vertical integration', in: E.J. Mitchell, ed., *Vertical integration in the oil industry*. Washington, D.C.: American Enterprise Institute.
Monteverde, K. and Teece, D.J. (1982a) 'Supplier switching costs and vertical integration in the automobile industry', *Bell Journal of Economics*, 13:206–213.
Monteverde, K. and Teece, D.J. (1982b) 'Appropriable rents and quasi-vertical integration', *Journal of Law and Economics*, 25:321–328.
Mueller, W.F. (1969) 'Public policy toward vertical mergers', in: J.F. Weston and S. Peltzman, eds., *Public policy toward mergers*. Pacific Palisades, California: Goodyear Publishing Company.
Oi, W.Y. (1961) 'The desirability of price instability under perfect competition', *Econometrica*, 29:58–64.
Ordover, J.A., Saloner, G. and Salop, S.C. (1987) 'Equilibrium vertical foreclosure', Department of Economics, New York University.
Parsons, D.O. and Ray, E. (1975) 'The United States Steel consolidation: The creation of market control', *Journal of Law and Economics*, 18:181–219.
Peck, M.J. (1961) *Competition in the aluminum industry 1945–1958*. Cambridge, Mass.: Harvard University Press.
Peck, M.J. and McGowan, J.J. (1967) 'Vertical integration in cement: A critical examination of the FTC Staff Report', *Antitrust Bulletin*, 12:505–531.
Peltzman, S. (1969) 'Public policy toward vertical mergers,' in: J.F. Weston and S. Peltzman, eds., *Public policy toward mergers*. Pacific Palisades, California: Goodyear Publishing Company.
Perry, M. (1986) 'An example of price formation in bilateral situations: A bargaining model with incomplete information', *Econometrica*, 54:313–321.
Perry, M.K. (1978a) 'Price discrimination and forward integration', *Bell Journal of Economics*, 9:209–217.
Perry, M.K. (1978b) 'Related market conditions and interindustrial mergers: Comment', *American Economic Review*, 68:221–224.
Perry, M.K. (1978c) 'Vertical integration: The monopsony case', *American Economic Review*, 68:561–570.
Perry, M.K. (1980) 'Forward integration by Alcoa: 1888–1930', *Journal of Industrial Economics*, 29:37–53.
Perry, M.K. (1982) 'Vertical integration by competitive firms: uncertainty and diversification', *Southern Economic Journal*, 49:201–208.
Perry, M.K. (1984) 'Vertical equilibrium in a competitive input market', *International Journal of Industrial Organization*, 2:159–170.
Perry, M.K. and Groff, R.H. (1982) 'Vertical integration and growth: An examination of the Stigler story', Bell Laboratories economic discussion paper no. 257.
Perry, M.K. and Groff, R.H. (1985) 'Resale price maintenance and forward integration into a monopolistically competitive industry', *Quarterly Journal of Economics*, 100:1293–1311.
Perry, M.K. and Porter, R.H. (1986) 'Resale price maintenance and exclusive territories in the presence of retail service externalities', Department of Economics, State University of New York at Stony Brook.
Pirenne, H. (1974) *Medieval cities: Their origins and the revival of trade*. Princeton, N.J.: Princeton University Press.
Porter, M.E. (1980) *Competitive strategy*. New York: Free Press.
Porter, M.E. and Spence, A.M. (1977) 'Vertical integration and differentiated inputs', discussion paper no. 576, Harvard Institute of Economic Research, Harvard University.
Porter, P. and Livesay, H.C. (1971) *Merchants and manufacturers*. Baltimore: Johns Hopkins Press.
Posner, R.A. (1976) *Antitrust law*. Chicago: University of Chicago Press.
Posner, R.A. and Easterbrook, F.H. (1981) *Antitrust*. St. Paul, Minn.: West Publishing.
Quirmbach, H.C. (1986a) 'Vertical integration: Scale distortions, partial integration, and the direction of price change', *Quarterly Journal of Economics*, 101:131–147.
Quirmbach, H.C. (1986b) 'The path of price changes in vertical integration', *Journal of Political Economy*, 94:1110–1119.
Riordan, M.H. (1984) 'Uncertainty, asymmetric information and bilateral contracts', *Review of Economic Studies*, 51:83–93.
Riordan, M.H. (1986) 'A note on optimal procurement contracts', *Information Economics and Policy*,

2:211–219.
Riordan, M.H. and Sappington, D.E.M. (1987) 'Information, incentives and organizational mode', *Quarterly Journal of Economics*, 102:243–263.
Riordan, M.H. and Williamson, O.E. (1985) 'Asset specificity and economic organization', *International Journal of Industrial Organization*, 3:365–378.
Salop, S.C. (1979) 'Monopolistic competition with outside goods', *Bell Journal of Economics*, 10:141–156.
Salop, S.C. and Scheffman, D.T. (1983) 'Raising rivals' costs', *American Economic Review*, 73:267–271.
Schmalensee, R. (1973) 'A note on the theory of vertical integration', *Journal of Political Economy*, 81:442–449.
Schmalensee, R. (1981) 'Output and welfare implications of monopolistic third-degree price discrimination', *American Economic Review*, 71:242–247.
Shavell, S. (1984) 'The design of contracts and remedies for breach', *Quarterly Journal of Economics*, 99:121–148.
Silver, M. (1984) *Enterprise and the scope of the firm*. Oxford, England: Martin Robertson and Company.
Smith, A. (1776) *The wealth of nations*. London: J.M. Dent and Sons, 1910.
Spence, A.M. (1976) 'Product selection, fixed costs, and monopolistic competition', *Review of Economic Studies*, 43:217–236.
Spengler, J.J. (1950) 'Vertical integration and antitrust policy', *Journal of Political Economy*, 53:347–352.
Spiller, P.T. (1985) 'On vertical mergers', *Journal of Law, Economics, and Organizations*, 1:285–312.
Stigler, G.J. (1951) 'The division of labor is limited by the extent of the market', *Journal of Political Economy*, 59:185–193.
Stuckey, J.A. (1983) *Vertical integration and joint ventures in the aluminum industry*. Cambridge, Mass.: Harvard University Press.
Sullivan, L.A. (1977) *Handbook of the law of antitrust*. St. Paul, Minn.: West Publishing.
Szenberg, M. (1971) *The economics of the Israeli diamond industry*. New York: Basic Books.
Teece, D.J. (1976) *Vertical integration and vertical divestiture in the U.S. oil industry*. Institute for Energy Studies, Stanford University.
Telser, L.G. (1979) 'A theory of monopoly of complementary goods', *Journal of Business*, 52:211–230.
Temin, P. (1987) 'Transactions costs and vertical integration: An historical test', *Journal of Economic History*, forthcoming.
Thorp, W. (1927) *The integration of manufacturing operation*. Washington, D.C.: U.S. Printing Office.
Tirole, J. (1986) 'Procurement and renegotiation', *Journal of Political Economy*, 94:235–259.
Tucker, I.B. and Wilder, R.P. (1977) 'Trends in vertical integration in the U.S. manufacturing sector', *Journal of Industrial Economics*, 26:81–94.
Varian, H.R. (1985) 'Price discrimination and social welfare', *American Economic Review*, 75:870–875.
Vassilakis, S. (1986) 'Increasing returns and strategic behavior', Ph.D. dissertation, Department of Economics, Johns Hopkins University.
Vernon, J. and Graham, D. (1971) 'Profitability of monopolization by vertical integration', *Journal of Political Economy*, 79:924–925.
Wallace, D.H. (1937) *Market control in the aluminum industry*. Cambridge, Mass.: Harvard University Press.
Warren-Boulton, F.R. (1974) 'Vertical control with variable proportions', *Journal of Political Economy*, 82:783–802.
Warren-Boulton, F.R. (1977) 'Vertical control by labor unions', *American Economic Review*, 67:309–322.
Warren-Boulton, F.R. (1978) *Vertical control of markets: Business and labor practices*. Cambridge, Mass.: Ballinger Publishing Company.
Waterson, M. (1980) 'Price-cost margins and successive market power', *Quarterly Journal of Economics*, 94:135–150.
Waterson, M. (1982) 'Vertical integration, variable proportions and oligopoly', *Economic Journal*, 92:129–144.
Weiss, A. (1987) 'Firm-specific physical capital: An empirical analysis of vertical mergers', Ph.D. dissertation, Department of Economics, University of Chicago.

Weitzman, M.L. (1974) 'Prices vs. quantities', *Review of Economic Studies*, 41:477–491.
Westfield, F.M. (1981) 'Vertical integration: Does product price rise or fall?', *American Economic Review*, 71:334–346.
Whalen, T.J., Jr. (1969) 'Vertical mergers in the concrete industry', *Antitrust Law and Economics Review*, 1:113–124.
White, J.J. and Summers, R.S. (1980) *Uniform commercial code*. St. Paul, Minn.: West Publishing.
White, L.J. (1971) *The automobile industry since 1945*. Cambridge, Mass.: Harvard University Press.
Wilk, D. (1968) 'Vertical integration in cement revisited: A comment on Peck and McGowan', *Antitrust Bulletin*, 13:619–647.
Williamson, O.E. (1971) 'The vertical integration of production: Market failure considerations', *American Economic Review*, 61:112–123.
Williamson, O.E. (1975) *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.
Williamson, O.E. (1979) 'Transaction cost economics: The governance of contractual relations', *Journal of Law and Economics*, 22:233–261.
Williamson, O.E. (1985) *The economic institutions of capitalism*. New York: Free Press.
Wu, S.Y. (1964) 'The effects of vertical integration on price and output', *Western Economic Review*, 2:117–133.
Young, A.A. (1928) 'Increasing returns and economic progress', *Economic Journal*, 38:527–542.