1 Introduction

In many network industries, most notably telecommunications, there has been a trend toward firms vertically separating by function, creating multiple, interdependent layers. For example, local broadband access networks consist of a conduit or poles layer, a cable coax or telephone copper wire layer, a powered channel layer linking customers to ISPs (where independence has been a contentious regulatory issue), an ISP-based Internet access portal layer, a content aggregators layer, and the content layer itself. We model this layering behavior, examining why it is happening, how it might impact competitive outcomes, and how it relates to the technology and the way it is implemented by operators.

We argue that vertical disintegration is motivated by economies of specialization. In most cases, these are associated with larger outputs and reflect technological and organizational characteristics akin to economies of scale. As the market grows larger and/or transaction costs of coordinating services across networks become smaller, these economies of specialization become relatively more important and lead to vertical disintegration. We show this outcome in a model that formalizes some of the insights in George Stigler’s 1951 article “The Division of Labor is Limited by the Extent of the Market.”

We consider whether incumbents can manage this process of vertical disintegration and entry. We ask whether the incumbent gains or loses from entry. We also examine whether the incumbent can

* CITI, Columbia University. E-mails: ad2239@columbia.edu and ch2287@columbia.edu
choose technologies that are socially inefficient (at least in the long run) to limit the scope of entry to complementary layers rather than the monopoly or oligopoly infrastructure. With sufficient technological lock-in, it may be possible for incumbents to limit the economies of specialization in ways that reduce the probability of entry into the infrastructure.

2 Stigler’s Model of Vertical Integration

Stigler published in 1951 a remarkable paper on Adam Smith’s famous theorem, “The division of labor is limited by the extent of the market.”\(^1\) Stigler’s paper is regularly cited in the economics literature even though it does not seem to have had very much impact on vertical integration theory. Stigler’s paper follows by 23 years Young’s (1928) pathbreaking paper on the same subject.

It is useful to analyze Stigler’s paper as consisting of three separate and largely independent stand-alone models. The first we will call the Division of Labor model, and it addresses Adam Smith’s conjecture. Stigler’s concern is based on the idea that the division of labor will be the source of firm-level economies of scale that would bring about a monopoly in each and every sector of the economy. As one does not observe monopolies throughout the economy, he asks whether Adam Smith’s conjecture is wrong and writes his famous paper to establish that, in fact, this not the case.

The second model we call the Process model. He looks at a firm as an organized set of separate and distinct processes, thus introducing a classical dimension to the neoclassical production function. He equates the conventional neoclassical cost functions on a one-to-one basis with the firm’s various individual activities. The approach is best conceived as an effort to bridge neoclassical analysis with team production (Becker and Murphy 1992). As shown by de Fontenay et al. (2004), it explains some flaws in treatments of economies of scale and scope (Panzar, 1989) as well as transaction costs (Williamson, 1985).

Stigler develops his Life Cycle Theory of vertical integration, what his paper is best known for, in the third stand-alone model. Describing firms in terms of a life cycle made up of three stages, he

\(^1\)A proposition Lowry (1879) traces back to the Oeconomicus by Xenophon. Kelly (1997) argues that the extent of the market itself is a function of factors such as the transportation and communication infrastructure that makes it possible to increase the extent of the market.
studies what may bring about vertical integration or disintegration in each of the transition periods between those phases. In the initial stage, vertical integration is inescapable because the firm has to carry out a number of activities that may not be available on the market. In the following stage, as the market matures, those activities are carried out by other firms that specialize in them. The outcome of the transition between those two stages is a gradual disintegration of the original firm. In the last transition phase, as the firm shrinks relative to the rest of the economy, vertically integration begins to reemerge. The Life Cycle model essentially reformulates the Process model in a quasi-dynamic form, using the dynamic structure to go around the static monopoly problem that was established in the first framework. This part of Stigler’s paper has been reviewed extensively and continues to receive the most attention. Stigler’s result is usually cited and set aside as it conflicts with the bulk of today’s vertical integration literature (Joskow 2005). We show here that the conflict may not be as real as apparent. Conventional analyses of vertical integration tend to be short run (Williamson, 1985), and, on occasion, medium run (Perry 1989), while Stigler’s framework is closer to Adam Smith’s long-run perspective, one that is more appropriate for policy formulation.

We argue that the individual models are most interesting as stand-alone studies, and this is what we focus on in this paper. We show that Stigler’s core issue – explaining why monopolies do not arise in Adam Smith’s framework – is not a problem in the first place. Adam Smith’s conjecture had already been established by Young (1928). Stigler’s concern reflects a confusion between economies of specialization associated with the division of labor and economies of scale that are central to modern theories of the firm. There are not necessarily links between the two, at least at the firm level, and for this reason monopoly need not arise.

3 Economies of Scale Versus Economies of Specialization

The concept of economies of scale (its foundations are discussed in de Fontenay et al. 2003 and de Fontenay et al. 2004) has always been problematic since it seems to lead inevitably to monopoly. Marshall (1920) stressed an aggregate concept of economies of scale with both internal and external economies rather than firm-level economies of scale. Later economists focused first on monopoly, then industry structure, and finally game theory to explain industries with market power.
In Stigler’s approach, the problem is pushed down from the firm level to the level of processes within the firm. He applies a conventional technology set (Mas-Colell et al. 1995) at a disaggregated level in the spirit of Adam Smith and his pin factory. The kind of processes Stigler lists to illustrate his approach includes processes such as “purchasing and storing materials, storing and selling outputs, extending credit to buyers” as well as the process of “transforming materials into semifinished products” (pg. 187). Effectively, Stigler’s firm is an entity that reflects an organized aggregation of distinct and separable activities that may correspond, say, to the units of its organization.

Stigler argues that the technological characteristics of those various processes are such that some may have substantial economies of scale. For instance, if we were considering a local telephone operator, most people think that the access network has substantial economies of scale while many of those same people are more at ease with the idea that those economies may not be particularly significant at the retail level. Stigler, just like Williamson (1985), does not see separability as a significant problem, a conclusion that is supported by some studies (e.g., Jacobides 2004) as well as by the high level of outsourcing one observes today (Feenstra 1998). Stigler observes that each of these cost functions will have their own technological characteristics, some, possibly, with significant scale economies, some, possibly, with diseconomies of scale. Thus, any economies of scale one may observe at the level of the firm are nothing more than ex post measures that need not describe in an ex ante manner the most efficient organization of the technology set.

This kind of approach was criticized in Young’s (1928) key contribution:

“the principal economies which manifest themselves in increasing returns are economies of capitalistic or roundabout methods of production . . . these economies lie under our eyes, but we may miss them if we try to make a large-scale production . . . as contrasted with large production any more than an incident in the general process by which increasing returns are secured and if accordingly we look too much at the individual firms . . . the economies of roundabout methods depend upon the extent of the market – and that, of course, is what we discuss under the head of increasing returns.”

Young argues that it is improper to equate economies of specialization created by the division of labor with neoclassical economies of scale. After a long hiatus, economists such as Yang (2001), Becker and Murphy (1992), Brown (1992), and Robertson and Alston (1992) are examining the
difference between the two. Those analyses show how the division of labor generates, as noted by Young, a downward sloping aggregate output curve, a curve that might be adequate at times at the industry level. But that curve does not tell us anything about the efficient firm size. A lower point along the curve could just as well be associated with smaller units of production as with larger ones. Stigler’s monopoly dilemma arose because he confused this \textit{ex post} residual curve for the economies of scale of a neoclassical production function. Yang (2001) develops this more formally:

“They system of production seems to exhibit economies of scale... But economies of specialization differ from economies of scale. First, economies of specialization are individual-specific and activity-specific... Second, the individual-specific time constraint and individual-specific system of multiple production functions are essential for defining the concept of economies of specialization.” (p. 46)

To incorporate these insights into a model, we adopt elements of both Stigler and Young, much as Yang (2001) does. From Stigler, we take the Division of Labor model, which could also be referred to as “Smithian specialization.” The key to that model is that if labor (or other resources) are allocated to a mix of two or more activities, they produce less in aggregate than if allocated to just one activity. From Young, we take the idea that these economies of specialization are only determinants of how factors should be allocated, not vertical integration.

Then what does determine vertical integration? Broadly, the various organizational systems that make up the “commons,” such as property rights, residual claims, contracting, governance, and so forth. Models of these systems include Alchian and Demsetz (1972), Grossman and Hart (1986), and others. We look at Coasian transactions costs as a convenient shorthand for these commons issues, and use them here just as in Yang. In fact, we interpret Williamson’s body of work as an extended discussion of why this shorthand works well. We acknowledge that transactions costs do not always equate to property rights based models (Whinston 2001), but we believe they serve well for our purposes.

4 Model

In this section we present a model that addresses the previous discussion. Following on that discussion, firms in our model do have economies of specialization, but they do not have economies
Thus, no firm can expand to create a monopoly, but there are incentives for firms to specialize and trade with one another. If there are limits on the number of firms that can enter (e.g. regulation or high fixed costs that create integer constraints), then the firms will earn Ricardian rents (or quasi-rents depending on the source of the entry barrier). It is these Ricardian rents, rather than oligopoly rents, that create incentives for firms to try to manipulate the market.

We begin with a simple case of vertical integration, and then consider the more complex case of specialization.

### 4.1 Vertical Integration

Firms produce a final good $q$. We can think of this as a consumer good, such as residential telephone service. Production of the final good is subject to decreasing returns to scale because of marketing and/or quality problems associated with large scale. Each firm is a perfect competitor that takes the price of $q$, $p_q$, as given.

Firms need two intermediate goods, $x$ and $y$, to produce output $q$. Let the production function be Cobb-Douglas:

$$q = f(x, y) = x^{1/4} y^{1/4}$$

Each firm has a quantity of critical resources $L$ available to allocate between producing intermediate goods $x$ and $y$. (Obviously the use of $L$ suggests that the critical resources are specialized labor, but in fact specialized capital is probably more relevant in network industries.) There are economies of specialization to producing either type of intermediate good: thus if the firm allocates $L_y$ resources to $y$ production and $L - L_y$ resources to $x$ production, its output is $y(L_y) = L_y^\beta$ and $x(L - L_y) = (L - L_y)^\beta$. The parameter $\beta > 1$ expresses the degree of economies of specialization. Using this notation, we can rewrite the production function for $q$ as

$$q = f(L, L_y) = \left((L - L_y)^\beta\right)^{1/4} \left(L_y^\beta\right)^{1/4}$$

The firm chooses $L_y$ optimally according to first order condition

$$\frac{\partial p_q f(L, L_y)}{\partial L_y} = \frac{\beta (L - L_y)^{\beta/4} L_y^{\beta/4 - 1} (2L_y - L) p_q}{4 (L_y - L)} = 0$$
which in this symmetric case gives a solution of $L_y = \frac{1}{2}$. The quantity of $q$ produced is

$$q_v^* = \frac{\sqrt{L^\beta}}{2^{\frac{d}{2}}}$$

Let demand be constant elasticity with $\epsilon = -1/2$. If there are $N$ firms, then the market price is

$$p_q^* = A(Nq_v^*)^{-1/2} = \frac{A2^{\frac{d}{4}}}{(L^\beta)^{\frac{1}{4}} \sqrt{N}}$$

At this price, the profits of one of the vertically integrated firms are

$$\pi_v = p_q^* q_v^* = \frac{A (L^\beta)^{\frac{1}{4}}}{2^{\frac{d}{4}} \sqrt{N}}$$

4.2 Specialization

Now we modify the above model to allow the firms to specialize in producing input $y$ or $x$. Assuming all firms still produce the final good $q$, specialization requires them to trade inputs with each other in order to satisfy the requirements of the production function.

Following Yang(2001), we add a “melting iceberg” transaction cost for factors traded between firms. For example, if a firm purchases 10 units of $y$, it may find that only 9 units actually contribute to production. The other unit (or at least the cost of it) can be thought of as melting away, representing a transaction cost. We use the parameter $k \in [0, 1]$ to represent the degree of transaction cost. In the example we just gave, $k = 0.9$. A higher value of $k$ means lower transaction cost, since more of the input is actually used in production.

4.2.1 Firm-Level Optimization

Consider a $y$ specialist. It uses all of its resources to produce $y$ ($L_y = L$). It sells some of the $y$ on the market ($y_s$) and uses the rest for production of $q$. It must buy the $x$ factor in the intermediate goods market ($x_d$). Then its production function is

$$q_y = f_y(x_d, L^\beta - y_s) = (k \times x_d)^{1/4} (L^\beta - y_s)^{1/4}$$

and its profit function is

$$\pi_y(q_y, x_d, y_s) = p_q q_y + p_y y_s - p_x x_d$$
Taking first order conditions, and assuming an interior solution, the $y$-specialist’s optimal purchases and sales in the intermediate good markets are:

$$x_d^* = \frac{\sqrt{kp_y^2}}{16p_x^\frac{3}{2}} \quad y_s^* = L^\beta - \frac{\sqrt{kp_y^2}}{16\sqrt{p_xp_y}}$$

Note an interesting effect here: an increase in $p_y$ causes the firm to increase $y_s$ production, which is not surprising, but it also causes a decrease in $x_d$ demand because there is a shift away from $q$ production.

For an $x$ specialist the problem is reversed but otherwise identical due to the symmetry assumption. An $x$ specialist’s optimal purchases and sales are:

$$y_d^* = \frac{\sqrt{kp_y^2}}{16\sqrt{p_xp_y}} \quad x_s^* = L^\beta - \frac{\sqrt{kp_y^2}}{16p_x^\frac{3}{2}}$$

### 4.2.2 Market Equilibrium

Let the number of firms of each type be $N_x$ and $N_y$. Then demand equals supply in the $y$ and $x$ markets requires that:

$$N_x y_d^* = N_y y_s^* \quad N_y x_d^* = N_x x_s^*$$

These can be solved simultaneously to give the equilibrium factor prices of $x$ and $y$ given the price of final output $p_q$:

$$p_y(p_q) = \frac{k^{\frac{1}{3}}N_y^{\frac{1}{3}}\sqrt{N_y + N_x p_q}}{42^{\frac{2}{3}}L^\frac{\beta}{3}N_y^{\frac{1}{3}}} \quad p_x(p_q) = \frac{k^{\frac{1}{3}}N_y^{\frac{1}{3}}\sqrt{N_y + N_x p_q}}{42^{\frac{2}{3}}L^\frac{\beta}{3}N_x^{\frac{1}{3}}}$$

So given $p_q$, the prices of $x$ and $y$ respond as one would expect to changes in the number of firms and hence in the demand or supply.

Now we need to find equilibrium in the $q$ (final good) market. The first step is to find out how much $q$ the two types of firms are producing. Substituting $p_y$ and $p_x$ back into the optimal purchases and sales allows us to find the factor usage of each type of firm, and thus its final good output. Because
of the symmetry of the problem, the final good outputs for the two types of firm are the same:

\[ q_x^* = q_y^* = \frac{2^{\frac{3}{4}} k^{\frac{3}{4}} L^{\frac{3}{4}} N_x^{\frac{1}{4}} N_y^{\frac{1}{4}}}{\sqrt{N_y + N_x}} \]

To understand this result, suppose that \( N_x > N_y \). Then the two types of firms still produce the same \( q \), but \( x \) firms use more \( x \) internally than \( y \) firms use of \( y \) internally. This is because \( y \) is more scarce and thus commands a higher price.

Now add up total supply and set it equal to demand. Then the equilibrium prices are:

\[ p_q^* = \frac{A}{2^{\frac{3}{4}} k^{\frac{3}{4}} L^{\frac{3}{4}} N_x^{\frac{3}{4}} N_y^{\frac{1}{4}} (N_y + N_x)^{\frac{1}{4}}} \]
\[ p_x^* = \frac{A k^{\frac{1}{8}} N_y^{\frac{3}{4}} (N_y + N_x)^{\frac{1}{4}}}{42^{\frac{3}{4}} L^{\frac{3}{4}} N_x^{\frac{3}{4}}} \]
\[ p_y^* = \frac{A k^{\frac{1}{8}} N_x^{\frac{3}{4}} (N_y + N_x)^{\frac{1}{4}}}{42^{\frac{3}{4}} L^{\frac{3}{4}} N_y^{\frac{3}{4}}} \]

Now we can find the profits of the two types of firms. They turn out to be rather complex expressions even for the symmetric case:

\[ \pi_x = p_q^* q_x + p_x^* x_x^* - p_y^* y_d = \frac{A k^{\frac{1}{8}} L^{\frac{3}{4}} N_x^{\frac{3}{4}} N_y^{\frac{1}{4}}}{42^{\frac{3}{4}} N_x^{\frac{3}{4}} (N_y + N_x)^{\frac{3}{4}}} + \frac{2^{\frac{3}{4}} k^{\frac{1}{8}} L^{\frac{3}{4}} N_x^{\frac{1}{4}} N_y^{\frac{3}{4}}}{2 (N_y + N_x)^{\frac{3}{4}}} + \frac{A k^{\frac{1}{8}} L^{\frac{3}{4}} N_y^{\frac{3}{4}} N_x^{\frac{1}{4}}}{42^{\frac{3}{4}} (N_y + N_x)^{\frac{3}{4}}} \]
\[ \pi_y = p_q^* q_y + p_y^* y_y^* - p_x^* x_x^* = \frac{A k^{\frac{1}{8}} L^{\frac{3}{4}} N_x^{\frac{3}{4}} N_y^{\frac{1}{4}}}{8 (N_y + N_x)^{\frac{3}{4}}} + \frac{2^{\frac{3}{4}} k^{\frac{1}{8}} L^{\frac{3}{4}} N_y^{\frac{1}{4}} N_x^{\frac{3}{4}}}{2 (N_y + N_x)^{\frac{3}{4}}} + \frac{A k^{\frac{1}{8}} L^{\frac{3}{4}} N_x^{\frac{3}{4}} N_y^{\frac{1}{4}}}{8 N_y^{\frac{3}{4}} (N_y + N_x)^{\frac{3}{4}}} \]

### 4.3 Comparison

Now the question that concerns us is whether a firm is better off in the vertical integration setting or the specialization setting. This depends on both the number of firms of each type (\( N_x \) and \( N_y \)) and the transaction costs involved in trading inputs between firms (\( k \)).

Due to the complexity of the profit functions, we will make the comparisons using a numerical simulation. Specifically, suppose \( \beta = 2, L = 2, A = 100, \) and initially \( N_x = 5, N_y = 5 \). For the vertically integrated case, let \( N = 10 \).

First, consider the effect of changes in \( k \). Profits of the two types of firms under specialization are the same due to symmetry. The results are in Table 1:
Profits rise as $k$ rises, indicating lower transactions costs. At about $k = 0.55$ it is more profitable to be specialized. This is due to the rather strong economies of specialization embedded in setting $\beta = 2$.

Now we repeat this exercise, but with $N_y = 4$ and $N_x = 6$. This introduces the idea that the $y$ firms are “incumbents” who may have some scarcity rents from their infrastructure. The $y$ factor will command a higher price, and the $x$ firms will be at a disadvantage. For the vertically integrated case, we report profits both for $N = 4$ and $N = 10$. The former is the case where under vertical integration, no $y$ is available to the $x$ firms and therefore they cannot produce. The latter is the case where all firms can produce $y$ under vertical integration.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\pi_v$ ($N = 4$)</th>
<th>$\pi_v$ ($N = 10$)</th>
<th>$\pi_x$</th>
<th>$\pi_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.4</td>
<td>50.00</td>
<td>31.62</td>
<td>28.41</td>
<td>33.19</td>
</tr>
<tr>
<td>.5</td>
<td>50.00</td>
<td>31.62</td>
<td>29.22</td>
<td>34.06</td>
</tr>
<tr>
<td>.6</td>
<td>50.00</td>
<td>31.62</td>
<td>29.89</td>
<td>34.84</td>
</tr>
<tr>
<td>.7</td>
<td>50.00</td>
<td>31.62</td>
<td>30.47</td>
<td>35.52</td>
</tr>
<tr>
<td>.8</td>
<td>50.00</td>
<td>31.62</td>
<td>30.98</td>
<td>36.12</td>
</tr>
<tr>
<td>.9</td>
<td>50.00</td>
<td>31.62</td>
<td>31.44</td>
<td>36.65</td>
</tr>
<tr>
<td>1.0</td>
<td>50.00</td>
<td>31.62</td>
<td>31.86</td>
<td>37.14</td>
</tr>
</tbody>
</table>

Table 2 ($N_x = 6, N_y = 4$)

This table gives several different results. First, for the $x$ firms, $k$ must rise to about 0.95 before
they prefer specialization to vertical integration (provided they could somehow produce $y$ under vertical integration). Second, the $y$ firms are indeed at an advantage, and they would rather move to specialization at very low $k$ values, at least for the case where there will be 10 firms in the market no matter what. This suggests some possible behavior by the $y$ firms to induce specialization of $x$ firms. Third, if the $y$ firms are the only firms that can ever produce $y$, then they are better off using that limit to keep the number of firms low. However, one can see that with different parameter values, there would be cases where the $y$ firms would prefer $x$ entry even under this scenario.

5 Conclusion

The model presented here shows how transactions costs can be used to proxy for the environment that controls vertical integration decisions. It also shows that when the number of firms is limited, different types of firms may have different incentives to move toward specialization.

We plan to extend the model in several ways. First, we will consider cases of incomplete specialization. For example, if there were four $y$ firms and one $x$ firm, then it seems likely that the $y$ firms would all produce a mixture of both factors, while the $x$ firm would probably specialize in $x$.

The next extension is to vary the factor intensity. As we suggested in the introduction, if the $y$ firms are first movers and can lock in the technology, then they may be able to manipulate future specialization decisions in their favor by changing the factor intensity. In fact, we expect that different factor intensities will be most profitable depending on the elasticity or inelasticity of demand for the final good.

Our final extension will add a $z$ factor which is an imperfect substitute for $y$. The $x$ firms will have the option of using this $z$ factor rather than buying $y$ from the incumbents. This will create a new switching point in the equilibrium, where there will be entry against the $y$ firms using this alternative infrastructure.
References


