A Stigler-Smith-Young Approach

to

Vertical Integration

Alain de Fontenay & Christiaan Hogendorn

1. Introduction

Stigler published in 1951 a remarkable paper on Adam Smith's (1776) famous proposition, "The division of labor is limited by the extent of the market." His paper is regularly cited in the economic literature even though it does not seem to have had very much impact on industrial organization and, especially, vertical integration (Perry 1989; Joskow 2005). His paper follows by 23 years Young's (1928) pathbreaking reformulation and informal proof of Smith's proposition.

The cornerstone of Stigler's effective framework is the division of labor, something that is both the newest and the oldest in the economic theory of production. Its reintroduction by Yang (2001), Becker and Murphy (1992), Fontenay and Hogendorn (2005) and others makes it the newest and this after almost a century during which it was mostly ignored by the profession. In a way, Stigler was not far from the truth when he was asserting in 1976 that "almost no one used or now uses the theory of the division of labor, for the excellent reason that there is scarcely such a theory" (p. 1209). Yet, it is also Stigler who, twenty five years earlier in 1951, had used Adam Smith's proposition, "the division of labor is limited by the extent of the market" as the title of his paper and the cornerstone of his proposed theory of vertical integration. The division of labor is also the oldest, with roots dating back at least to Xenophon and his shoe factory in the fourth century BC.

Cournot in 1838 was the first to face the problem Stigler identifies with Smith's proposition, namely whether increasing returns is compatible with competition. Marshall faced the same problem; White (2002) depicts vividly the environment Marshall was trying to apprehend when he confronted the same perennial problem. White tells us of producers who "huddle together as an industry such that their key cues come from their fellow producers who face the same opaque diversity of buyers." It is on that dimension, the inherent interdependence of firms competing with one another in a world of uncertainties, rather than Cournot's world of imperfect competition that attracted Marshall's attention. It is that world of "competition" that led him to conceive the externalities. That solution was hardly acceptable to a new generation of economists, Knight first among them, who were more attracted by the neoclassical model's mathematical elegance. Knight (1925) argued that he could not conceive of Marshall's externalities. Sraffa (1926) followed by others including Stigler in 1951 saw Marshall's externalities as a "cop out." Sraffa formalized the incompatibility of increasing returns with perfect competition, providing Stigler with his entry in the 1951 paper. Even though that result became the cornerstone of neoclassical analysis, Chipman (1970) was able to establish restrictive neoclassical conditions when it did not hold.

Stigler chose an original approach to the theory of the firm. The Stiglerian firm, just like the Coasian firm, is made up of a set activities. It is also true that, as with the Coasian firm, the activities that are internalized within the firm because they lower the firm's costs. However, Stigler was looking to go further. He was interested in explaining why a specific activity might be integrated in the firm's set of activities, not just on the basis of transaction costs but also in terms of the characteristics of the activities. Stigler situated his firm in time; the firm starts as a new entrant and innovator and grows to maturity through the successful implementation of that innovation.

Because the firm's entry is based on innovation, there are intermediate products the firm needs that are not available on any market. As the firm grows, it attracts somehow other firms resulting eventually in a new, competitive industry. Some of the activities that are internal to that industry attract entrepreneurs who, in turn, put the entry cycle in a new phase. Stigler's

model has its roots in Smith's division of labor. For instance, Stigler, just like Smith and Coase, relies on competition among firms to impose efficiency through market discipline. We are largely back to Smith's division of labor model.

Unfortunately, Stigler does not take his model that far. His framework is neoclassical, based on a set of activity-specific production functions. In addition, he insures that he does not deviate too far from a Knightian model by imposing additional constraints upon his model's dynamic structure. Effectively, he assumes that the firm goes through a simple life cycle. As a result, he falls short from his goals. Nevertheless, the paper makes significant contributions and those have not been properly recognized. For instance, by positioning the integration and/or disintegration of activities in the perspective of innovations that go through their own life cycles, he is introducing de facto a dimension that is still not recognized. Nevertheless, Stigler's choice is an acknowledgement of the role of innovation within vertical integration. Stigler's model implies an innovation-based perspective that differs from the conventional treatment of vertical integration.

Before looking at the neoclassical perspective, it is useful to situate the dynamic dimension of the neoclassical analysis in its proper perspective. Knight, like Marshall and Young, recognized that monopolies do arise but, like Marshall and Stigler, he also argued that they are uncommon. Then he argues that monopolies, to the extent they arise, will be found in small but expanding industries. His justification for that condition is that industries, as they are expanding, always face pressure from increasing costs. Those could be factors such as the "internal costs of cohesion in a human organization" (Knight 1926, p. 56). In any case, he argues that eventually, and sooner than later, "economies of organization are offset by special costs due to size" (Knight 1924, p. 366). Knight is also, implicitly, concerned with internalized externalities. To eliminate the danger, he argues that an industry must include "the entire process of producing a *finished product*" (Knight 1926, p. 56).

Our objective in this paper is to contribute to the history that preceded Stigler's (1951) paper. However, we have another objective. The basic problem with increasing returns and competition does not affect Smith's analysis. It is a problem with the static nature of any neoclassical framework. Stigler's exploration of the subject highlights to what extent the classical division of labor can be a natural foundation for the analysis of both vertical integration and innovation. Here, we set aside the neoclassical path to reset the Stiglerian firm in the Smith-Young's classical division of labor framework. By this step, we are able to outline a powerful vertical integration framework that can be used to develop a new theory.

The classical division of labor provides us with a dynamic where firms innovate to survive and gain some control over their market and innovate to achieve some stability vis à vis their competitors (Fligstein 2001). Firms, as pointed by Marshall are not all identical. Some survives from having been on the market for a long time even if they may not all be that efficient. They may be able to survive from the inertia of their customers and from the established relationship they have with them. Other firms may be dynamic, making effective use of successful innovative steps at any of a number of different stages in their production process. Such firms have an incentive to try to gain control of the underlying technology the way firms often do it in infrastructure sector such as telecommunications (de Fontenay, Liebenau, and Savin 2004) in order to make entry costly to competitors, hence to gain and eventually sustain market power. Then there are firms that are new entrants, firms that try to challenge the established order or to establish new sectors as in the telecommunications wireless sector.

Stigler's dynamic analysis is based on the firms' life cycle. The innovation dimension that is inherent to the division of labor suggests that there may be a different way to look at the life cycle process. In Stigler's approach, the innovation is equated with the firm and his dynamic model could be looked at as a dynamic treatment of innovation. However, had he done that, he would have had a stronger ability to model the evolution of the firm through time and its incentive to control vertical integration and disintegration. A life cycle model of innovation would have taken him much closer to the theory of vertical integration he was trying to create.

We can start to see what the element of a theory of vertical integration might look like. The technological and organizational framework is the historical division of labor. Smith's contribution was to realize that the extent of the market was more than a body count, i.e., that it involved also some measure of purchasing power. It was also to introduce competition among producers. To take fully advantage of the division of labor, one has to embed it in a dynamic context. Smith used the interaction between the division of labor as innovation and competition among producers as the dynamic motor that produce growth and wealth. Once innovation and competition are intertwined as interdependent dynamic processes, then an innovation life cycle model reinforce the time dimension of the dynamic framework.

Those elements provide an infrastructure that would seem adequate to build a theory of vertical integration. However, they would also seem to favor disruptive innovation, weighing toward more instability than what one would expect in Smithian competition. The Smith-Marshall-Young concept of growth competition in an environment where innovations tend to emerge from the division of labor is one that would get its consistency from a mix of sustaining and disruptive innovations. Stigler's model would be one where disruptive innovations dominate.

White (2002) and especially Young (1928) stress that Marshall's externalities, his external economies at the level of the industry, are his key insight. Young argues convincingly that the external economies spread beyond the firm and the industry to other segments of the economy. External economies have been ignored if not systematically attacked by economists. We can site Knight (1925), Sraffa (1926), Stigler (1951), Liebowitz (1995), and Haddock (2004). There are only a few economists that have looked at them seriously (Buchanan and Stubblebine 1962; Rosenberg 1963; Strassmann 1959 and 1963; Chipman 1970; White 2002).

Economists tend to downplay externalities because they seem reluctant to deal with the internalization problem. Knight's (1924) often cited quote is seen by many as a warning: "External economies in one business unit are internal economies in some other, within the industry" (p. 597). Knight is evidently right within the strict confines of neoclassical analysis but this is of little relevance since even the smallest marginal deviations such as information asymmetries can easily upset the neoclassical balance.

It is Young (1928) that has been the most successful in developing and formulating an unambiguous and clear concept of externalities and to express the cost to economic analysis when they are neglected:

> "[T]he 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 [one of the form of division of labor] depend upon the extent of the market – and that, of course, is what we discuss under the head of increasing returns."

It is also evident that Young poses an enormous challenge to formal analysis because he has nothing less in mind than a full, dynamic, largely disequilibrium vision of the economy.

Interestingly enough, the production (or cost) function is the source of the problems economists face with those external economies. Adam Smith was using the terminology of "increasing returns" just like Knight, Stigler, and Young and just like us today. Classical increasing returns based on the division of labor became neoclassical increasing returns based on production functions. The transition went by without anyone asking whether the two were the same.

In this paper, we show that the division of labor does not correspond to the production function approach. Rather, we show that, in the simplest form of division of labor, the production

process is achieved in two distinct steps. In the second process, new inputs in the form of newly specialized labor with new, different machines are transformed through a new production function into the desired output, whether the latter is an intermediate or a final good. The neoclassical production function effectively aggregates that production function with another separate process in which the former inputs are transformed through education, training, innovation, and other processes into the new inputs (Barnett 1925). The division of labor tells us that the transformation of the inputs is characterized by substantial economies of scale since this is the only reason for going through such a transformation. The division of labor only tells us that the aggregate of the two sets of transformations must also be characterized by scale economies. This condition does not constrain the technology of the division of labor's production process, hence, the classical production function, to also be characterized by scale economies. It could easily be an increasing cost process. This is why Adam Smith did not have to worry about monopolies when considering whether the division of labor is limited by the extent of the market. This why Marshall was on the right track with his external economies, notwithstanding Knight's arguments to the contrary and this is why Stigler's criticism of Young is not founded. This is why, finally, we are able to merge elements of Stigler's analysis within Young's framework to create a conceptually sound approach to vertical integration.

2. Stigler's paper

2.1. Introduction

Stigler suggests that Smith's proposition, "the division of labor is limited by the extent of the market," creates a dilemma that he calls a "superficial dilemma." Smith's proposition means a monopoly in every sector of the economy, i.e., a market structure that is inconsistent with everyday observation (White 2002), hence a dilemma. Stigler's dilemma must presumably be based upon Sraffa (1926) since Stigler's objective is to establish Adam Smith's proposition within a "stable competitive equilibrium." The benchmark he sets for his review of the literature is a solution that provides a stable competitive general equilibrium. On that basis, he concludes that not even Young (1928) had been able to "resolve the technical difficulties of incorporating the extent of the market into competitive price theory" (p. 187).

While Stigler introduces his analysis with a "superficial dilemma," the monopoly problem conflicts with both everyday observation and Adam Smith's arguments in support of the proposition. Firms are implementing the division of labor as a way to gain a comparative advantage over their competitors. The competition among firms suggests that one of those firms should in time outdo its competitors through greater and more efficient division of labor. That firm should become a monopoly in time. In other words, the proposition must mean that every sector has to turn into a monopoly.

Stigler's 1976 comment suggests that he came to that conclusion because he did not see any real difference between increasing returns and economies of scale. Then, the division of labor implies increasing returns that mean, in turn, by Stigler's assumption, economies of scale that are incompatible with competition. Stigler's next step consists in proposing a new conception of the firm that is inspired by the division of labor even if it is neoclassical in nature and construction. However, representing the firm as an aggregation of elementary activities à la division of labor does not do anything to address the monopoly problem. Sraffa's (1926) analysis eliminates the possibility of a static solution to the economies-of-scale-based monopoly problem.

Predictably, Stigler looks for a solution in transforming the problem by making it dynamic. He selects a simple but artificial model, the life cycle model of the industry and, by extension, the firm presumably for its simplicity. Yet, this solution is incompatible with his competitive general equilibrium objective and it can hardly be looked at as the building block for a "stable equilibrium."

The life cycle model Stigler adopts is little more than an ex post, superficial cover for his dynamic analysis that is inherently classical and inherently an evolutionary analysis. In that analysis, firms start somehow as monopolies to give way by some unspecified process to a competitive industry. It is easy to see that Stigler has not really progress beyond Young's (1928) work, i.e., that Stigler has failed to achieve his stated goal. Yet Stigler's paper is a staple of the vertical integration literature (Perry 1989) that is regularly cited even if it is just as regularly criticized (Williamson 1975; Levy 1983) if not dismissed (Joskow 2005).

We conclude that Stigler did not achieve any of his objectives. We conclude also that Stigler made nevertheless a substantial contribution. He gave impetus to the growing new classical literature (Borland and Yang 1989; Becker and Murphy 1992; Fontenay and Hoggendorn 2005)

Stigler's analysis is best viewed as a three-legged stool. Each leg corresponds to a distinct and largely independent model that is intended to meet one of Stigler's objectives. With the first stool Stigler rephrases the neoclassical monopoly problem within the classical division of labor framework and introduces the firm via the monopoly. The second leg consists in a neoclassical representation of the classical firm, say, Adam Smith's pin factory, or, preferably, one of its modern reincarnation by Babbage (1832) or, even better, by Young (1928). The third leg is a dynamic model of the firm, as developed in the second model. Stigler implicitly presents the

firm as belonging to an industry both of them evolving through a simple life cycle.

While Stigler's treatment is extremely simplified, industrial organization would benefit from adopting it systematically. For instance, as shown by Fontenay et al. (2004), it establishes some of the fundamental flaws in today's treatment of economies of scale and scope (Panzar, 1989) as well as in today's treatment of transaction costs (Williamson, 1985). At the same time, it raises questions about existing technology or transaction cost-based theories of the firm.

At that stage, Stigler notes that each activity is described as a conventional neoclassical firm, with its own production function. The production functions of the various activities could be expected to differ from one another in terms of their characteristics. This means that some may have constant returns to scale, some increasing returns to scale, some decreasing returns to scale, and others with more complex structures. This means that, when discussing the economies of scale of the original firm, those economies of scale reflect activity-specific economies of scale that are concentrated in only a few activities. In other words, the firm's monopoly problem is not a problem at the level of the firm but a problem at a far more disaggregated level.

Stigler's analysis is not surprising. If one think about local telephony, it is surprising that people so often argue that they are characterized by large economies of scale. Stigler's approach highlights the fact that those economies of scale are not evenly distributed across all the local activities. The real estate, i.e., the trenches that were required to build ducts and conduits and the poles that are used for copper pairs, coax cables, and fiber, represent the main source of economies of scale.

Stigler's originality is to have cast the problem unambiguously in a quasi-classical framework with its roots in Adam Smith's proposition. Cournot, Marshall, and Sraffa were already working with neoclassical-type production functions that were equated with the firm. Stigler's production functions correspond to individual activities carried within firm, very much along the lines of Adam Smith's famous pin factory example. That way, he is the first to model the classical division of labor.

2.2. Adam Smith's proposition, Stigler, and the monopoly

Stigler's analysis is very critical of both Marshall and Young and, yet, he does not contribute any new insight beyond the generic neoclassical criticism. He dismisses Young (1928), describing his key contribution as an effort to rescue the concept of economies of scale through an endorsement of Adam Smith's proposition. He adds that Young was only able to go part of the way in addressing the monopoly dilemma Stigler saw at the heart of Adam Smith's proposition.

Once one accepts that economies of scale is just an alternative to the division of labor's increasing returns, then Stigler's argument that Smith's proposition implies a monopoly in each and every sector is nothing more than a restatement of Sraffa (1926). More significantly, thus recast, Adam Smith's proposition cannot be consistent with competition. Let's consider Xenophon's shoe factory, Athens, and some village in the hinterland of Greece or Macedony. It is reasonable to treat Athens and the village as distinct markets. If this is the case, the division of labor tells us that there must be one person, working part time who was making shoes in the village. That individual's shoe enterprise, no matter how small, had to be a monopoly and the price that individual charged to the villagers had to reflect the villagers' demand elasticities. The situation could not have been different in Athens's shoe market. There, one manufacturer had to be more successful than the others and take over the manufacturing of shoes for Athens and the extent of the market for shoes around Athens. In other words, the division of labor meant that in Athens as in the village, shoes had to be produced by a monopoly and that monopoly's pricing had to be limited by elasticity of demand for shoes.

What is the way around the monopoly's problem? Stigler does not offer the appropriate answer within the scope of his stated goal: "It is the central thesis of this paper that the theorem of Adam Smith… is the core of a theory of the functions of firm and industry…" (p. 185).

Effectively, he assumes that innovations are bringing about new sectors that impact, if we were to stay in Xenophon's context, the shoe business, but this is not sufficient. We know from our experience with sectors such as telecommunications that innovation per se is not likely in general to bring about competition. One has to make additional assumptions, say, that due to some other constraints, say, limitations in how efficient management is, after a certain size, the firm's economies of scale give way to diseconomies of scale.

This has to be the kind of hypothesis Stigler has in mind when he argues that, with market expansion, the monopoly would give way to competition. But this is nothing more than to argue, within Sraffa's context, that the economies of scale are limited to a range of output that is sufficiently small that no one firm has significant market power. It implies that it is not the extent of the market that is limiting the division of labor.

2.3. Classical vs. neoclassical analyses

The neoclassical treatment of the monopoly (Knight 1925), Sraffa's criticism of Marshall's externalities, and Stigler's (1951) dismissal of Young's (1928) treatment of Adam Smith's proposition are disingenuous. Both the classical division of labor and the neoclassical production or cost function are frameworks to model production processes in the economy. However, it is important to understand where they stand with respect to one another. The production function is a static tool that describes the production process at a point in time. The division of labor, on the other hand, describes the technology as it evolves. While we might use comparative static to compare a firm at two different output levels, the analysis does not tell us how the production process would evolve to go from one point to the other. It only tells us what the firm's output would be at interim points. The division of labor on the other hand describes what makes the production process evolve through time. The division of labor helps us infer the kinds of resources the firm might require to produce some output level.

Let's use a simple analogy to stress the nature of those differences. Zeno's paradox challenged thinkers over many centuries. If you fire an arrow with a bow and if you look at the arrow, during its flight at a given point, you have an arrow that is somewhere in the air. Zeno asked why is it that the arrow falls straight down as a result of gravity. We know that the paradox is resolved because Zeno's description of the arrow at that selected point does not include other forces that are affecting the arrow's flight path. Applying Zeno's paradox to the division of labor and the neoclassical production function, we see that the latter corresponds to Zeno's description of the arrow as, during its flight, it is at a certain point in the air. On the other hand, the classical analysis of the division of labor provides us a description of the process that got production at the observed production level and that will take it further. In other words, the production function is a static and, by necessity, incomplete description of the production process to the extent that it does not tell us, from an ex ante perspective why production is where it is and where it can be expected to go after. This is what the division of labor and more generally classical economics studies.

The neoclassical criticism of the division of labor and its extension such as Marshall's externalities whether it is by Sraffa or Stigler is disingenuous. The significance of the neoclassical contribution was twofold, first to place much greater emphasis upon the individual agents in the economy and then to impose a degree of rigor that was largely lacking in classical analysis. There is absolutely no reason to imagine either Babbage (1832) or Marshall (1890) would have had problems with the neoclassical monopoly argument. They would simply have found it irrelevant and of no significance.

The new endogenous growth theory (Romer 1986, 1990) is the first effort made to impose rigor to Young's analysis and, still, it does it by imposing artificial constraints on Young's work, reducing his model to an equilibrium model.

2.4. The Stiglerian firm

Stigler's greatest contribution in this paper is his model of the firm. The Stiglerian firm is

complementary to the Coasian firm. The Coasian firm is defined in terms of the relationship between internal and external transaction costs in such a way that where internal transaction costs across a set of activities are lower than the external transaction costs, then the activities will be organized as a firm. Stigler shows some awareness of transaction costs but his principal focus is on the technological characteristics of the various activities and on their evolution.

Stigler's analysis is carried out at times in terms of the industry, at times in terms of the firm. The industry does not play any role in Stigler's analysis other than as a marker to denote competition: "The firms will… abandon the process (Y_1) and a new firm will take it over. This new firm will be a monopoly… With the continued expansion of the industry [the set of firms that was producing Y_1 as one of the intermediate products], the number of firms supplying process Y_1 will increase so that the new industry becomes competitive."

The Stiglerian firm is specified in terms of a set of activities such that each of these activities can be described by a conventional production (or cost) function, implying that it could be a stand-alone firm. 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" (187). What concerns him is the technologies for such well-defined activities as billing,… because those can be more meaningfully associated with coherent and integrated individual processes. Stigler does not impose any constraint on those individual processes except to argue that they will be typically described by technologies that are likely to differ systematically from activity to activity. Stigler notes that the scale characteristics of those production functions will vary across activities.

Rosenberg and Ames () provide a set of definitions that help better appreciate the static form of the Stiglerian firm: (1) commodity or, oversimplifying, output; (2) factor or, oversimplifying again, input; (3) activity, or in Stigler's terminology, function; (4) process; (5) firm. To stress the vertical structure and the potential stand-alone capability of individual activities, it is important to consider activities as the activities' inputs that will generally be a mix of inputs and intermediate products. i.e., products that are produced by upstream activities

If we take Adam Smith's model of the pin factory as the model of a classical firm, we would have a firm that carries a number of functions or activities the way Adam Smith described them for the pin factory. Each of these functions could be completed by a set of specialized workers and each function, under adequate organizational and transaction costs conditions, could be conceived as a stand-alone firm. This is the way Stigler approaches the modern firm. Taking a neoclassical perspective, each function, say, human resources, could be viewed potentially as a stand-alone firm with its own production and cost functions. Stigler considered only application of the division of labor at the firm level, a novel approach at the time.

The Stiglerian firm's production activity consists effectively of the aggregation of a finite and exhaustive set of production processes. The level of analysis he adopts is still highly aggregated and characterized by a high level of abstraction relative to Young's (1928) "economies of capitalistic or roundabout methods of production," hence to the division of labor (Yang 2001). Nevertheless, it is far more descriptive and "classical," hence, differs in fundamental ways with the conventional cost function (Panzar 1989) and the conventional corresponding economies of scale (Williamson 1985).

It is reasonable to assume that Stigler would have no problem with Williamson's stress on transaction costs and opportunism as long as those are constrained to adjustment problems that might delay disintegration, i.e., would be endogenous and sensitive to learning-by-doing. It seems fair to argue that this is the way he would look at dimensions such as bounded rationality, opportunism, or asset specificity. However, that would not be the case. Williamson has not acknowledged the time dimension of his analysis and, for that limitation in his framework, the two would eventually have to strongly disagree. Stigler fully shares with Adam Smith the role of technology in shaping the market environment the way Adam Smith had done

and he could not accept to jettison it to accommodate Williamson. Conversely, from Williamson's perspective, Smith's proposition is essentially irrelevant, hence of little interest, a perspective that we find again in Becker and Murphy (1992) even though their work is framed from a pure Chicago perspective.

The time dimension, i.e., whether Adam Smith's analysis holds in the short run or is a long run argument, is the fundamental difference between Stigler's and Williamson's perspectives. Stigler would argue that the various factors Williamson identifies are certainly relevant within a short-to medium-run perspective. He would then point out that their significance would progressively vane and eventually become so minute as to be irrelevant. His conclusion would then be that they do not affect Adam Smith's proposition. Furthermore, the kind of attack Stigler would pursue would be based on the very instruments Williamson identifies to argue the lack of relevance, in today's world, of the division of labor.

After all, other things equal, time would help stakeholders to learn and improve their ability to cope with bounded rationality. In other words, through time, from his perspective, people would find ways to address more and more efficiently challenges such as bounded rationality. In other words, the hold of those issues in preventing the extent of the market to become a constraint on the division of labor would weaken. Sengupta (2001) is one of those who have demonstrated the ingenuity of human societies in finding solutions in the form of managed commons, using the expression coined by Hardin (1998) to the kind of challenges Williamson use to build his system.

One could argue that governance that are able to tackle factors such as opportunism or asset specificity are not trivial and that firms as any other forms of communities may not be able to always resolve them – Bowles opens his 2004 book with an example of such an opportunism-based governance failure. Nevertheless, in view of the strong incentive they have to find reasonable ways to deal with opportunism, one needs to conclude that firms would tend to be successful, at least in the longer run neglected by Williamson, to resolve those problems (Coase 2000; Bowles 2004), an argument found in Adam Smith (Rosenberg, 1976).

To conclude our analysis of Williamson's critique of Stigler and of his bias in favor of short run outcomes, it suffices to point to Williamson's (1975) average cost curve analysis pp. 17-18. There he differentiates established firms with new entrants, stressing that the fixed entry costs a new firm would have to cover have already been covered by the established firm. In the p. 18 diagram, Williamson sets the average cost curve for a new firm everywhere higher than those of the established firms on the same ground. We know from conventional economic analysis that firms need not consider their sunk costs when they price their services, but that argument is only meaningful in the short-run. Long-run pricing has to look at sunk cost as variable and a factor that affects their cost of capital. To stay in business, the firm has to look at its long run interests and its long run curve would be set in terms of what an entrant is willing to invest to enter. Our point is that here also Williamson has confined his arguments to the short-run, i.e., that he is not addressing the problem that concerns economists such as Adam Smith, Young, and Stigler.

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, 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 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 argues that each of these cost functions will have its own technological characteristics, some, possibly, with significant scale economies, some, possibly, with diseconomies of scale,…

Observing that "[c]ertain processes are subject to increasing returns," Stigler asks the question, "why does the firm not exploit them further and in the process become a monopoly?" Williamson (1975) sees the transaction costs as the self-evident response to Stigler's question but that answer is disingenuous since the production function incorporates all elements, including transaction costs as expressly noted by Stigler.

The question Stigler asks is very much the question raised by Knight about the monopoly as cited by Coase – just as Williamson raises an orthogonal issue while failing to address Stigler's question, Coase does the same thing vis à vis Knight. Effectively, Williamson, duplicating Coase vis à vis Knight, ignores the economies of scale problem. It is true that the transaction cost dimension is treated in far greater depth by Williamson than by Stigler who only demonstrates some vague awareness of the problem.

Stigler bypasses the monopoly problem he acknowledges. Where the product is so new that there does not exist a market, then he observes that the firm will produce it internally. Why talk of new products? He describes the early steps of innovation the way Chandler would do it: "Young industries are often strangers to the established economic system. They require new kinds or qualities of materials and hence make their own… These young industries must design their own specialized equipment, and often manufacture it… When the industry has attained a certain size and prospects, many of these tasks are sufficiently important to be turned over to specialists…"

We see that Stigler's vertical integration is different from Coasian vertical integration. Stigler describes an emerging sector with emerging firms that are innovators and are trying to transform themselves in order not to be strangers to the economy. Stiglerian firms, at that stage, are only vertically integrated from an ex post perspective, once 'specialists' appear. This vertical integration is not the vertical integration within a static neoclassical context, the way Coase studied the subject in 1937. It has nothing to do with the vertical integration one observed in sectors such as telecommunications. Stigler's analysis has more to do, at this stage, with a theory of innovation as with the conventional literature on vertical integration (Perry 1989). Interestingly enough, it is a study of the firm and of its organizational structure that is entirely compatible with the division of labor, especially as developed by Young (1928).

2.5. Other reviews of Stigler's 1951 paper

2.5.1. Introduction

Perry (1989) highlights the inconsistency between Stigler's (1951) theory and today's dominant view on vertical integration, namely that vertical integration increases as the firm expands and become more established. Many such as Joskow (2005) look at Stigler's approach as exclusively rooted in an overly narrow framework, economies of scale within a the technology theory of the firm.

New institutional economists point to a growing body of evidence to stress that organizational and transaction costs as well as information, education, and coordination costs play a fundamental role as a source of constraints on the division of labor. Becker and Murphy (1992) go as far as arguing that organizational and transaction costs as well as information, education, and coordination costs may be so important that the extent of the market might become irrelevant. Joskow (2005), just like Williamson (1975), focuses essentially the theory of vertical integration upon a transaction costs-based analysis of the firm. Those show little interest in neoclassical market failure-based analyses of vertical integration. As far as Stigler's (1951) vertical integration model, they point to the lack of empirical evidence.

Yet, Brown (1992) finds considerable support for the sensitivity of vertical integration to the extent of the market in his comparative study of the textile sector in Germany and the U.K. in the pre-WW I period. In his model, the variation in the extent of the market is traced to the protection against international competition, protection that simultaneously restricts the market and creates incentives for less disintegration where there is more protection.

Interestingly enough, Brown's result conflicts with Elberfeld's (2002) analysis that shows that Stigler's vertical integration model holds within an equilibrium model provided there is no restriction upon competition.

This highlights the complexity of the vertical integration problem. Stigler recognized the significance of organizational and transaction costs as well as information, education, and coordination costs in vertical integration. Such costs are hard to dissociate from costs that are implicit in the division of labor process. Those costs are certainly intuitive and yet their specification and the way to assess them is a major challenge as demonstrated by Coase (2000). In his 2000 paper, Coase used his direct experience while visiting Detroit in the early thirties and his interviews on the matter of the acquisition of Fisher Body by General Motors to criticize the empirical foundation of the paper that has become the classic on vertical integration, (Klein, Crawford, and Alchian 1978; Klein 1988; Freeland 2000).

Far more important, those costs are also very time sensitive. For instance, factors such as learning processes and the development of routines contribute to the lowering of costs from period to period. Learning and the development of routines reflect the unique human ability to adjust and identify innovative ways to improve existing processes – we have already discussed the example Adam Smith gives of a young boy looking for ways to shirk the drudgery of his factory work. People seem to forget that the role of transaction costs had already been acknowledged by Stigler in 1951 as when he suggests that "the sale of the product may be too small to support a specialized merchant " (188).

The economic literature does not challenge Stigler on the question of economies of scale as, implicitly, an appropriate measure for the economies of specialization the division of labor generate. Demsetz (1988) and Becker and Murphy (1992) uses those economies of scale in their own core analyses while they are implicit but not a subject of investigation for new institutionalists as illustrated by Williamson's (1985) heuristic model of vertical integration.

While there is little doubt that Perry's (1989) overall conclusion still holds as demonstrated by Joskow (2005), there is a literature that is broadly consistent with Stigler's model (Christensen, Verlinden, and Westerman 2002; Arora, Bokhari, and Morel 2005; Arora and Bokhari 2005). Interestingly enough that recent literature, just like Stigler, gives a much greater role to innovation.

2.5.2. Demsetz

Demsetz (1988) provides a clear, concise overview of Stigler's paper, where he identifies the problem as a conventional vertical integration problem, i.e., the pure "build-or-buy" problem. Hence, just like Williamson, he reflects concerns that, we would argue, were already sketched by Stigler, say, when he asks, p. 187, whether "the entrepreneur must neglect production in order to supervise marketing." Predictably, this is Demsetz's Achilles heel.

Demsetz finds Stigler's treatment of the specialization problem overly restrictive. He points out that Stigler's model treats only one among a large set of options one would want to evaluate. He argues that those other options alternatives to Stigler's vertical disintegration that are often observed in practice. One of these options would be for the firm to reorganize creating new departments capable of responding to the conditions the division of labor could be creating internally.

Demsetz argues also that, through expansion, firms will eventually become sufficiently large to be able to take advantage of all economies of scale internally. This will happen once they have reach a size where all scale economies are exhausted internally. Beyond that stage, the impact of the division of labor will remain internal to the firm and the specialization of firm would not tend to increase anymore. Demsetz adds that there may be other effects that may make the situation even more muddy. For instance, the growth in the extent of the market may impact the price of some of the factors inputs. In conclusion, he notes that there are so many factors at play that it is hard to come to any clear conclusions even though he is ready to accept

Stigler's solution as a likely outcome.

Demsetz does not reject the division of labor or technology-based treatment of the firm in principle the way Williamson does. Nevertheless, there are still flaws in his analysis. He considers that specialization needs not automatically result in the disintegration of the firm and in the creation of a new firm specializing in the activity that is characterized by substantial scale economies. Stigler would not necessarily challenge that this may be the case for any of a number of activities.

The point is that there is no natural boundary to how far one could dream of a vertical disintegration. In practice, it is the market, especially, the willingness of entrepreneurs to innovate and create new activities and searching for more efficient solutions that gives the division of labor its dynamism. The environment and many other factors would impact how far, at anyone time, the specialization would go. The process would also clarify the path the disintegration would follow. Rather, this is the very essence of the Adam Smith proposition Stigler is trying to establish.

Sigler's inquiry does not preclude that one would observe a trend toward specialization as the extent of the market expands. It is also possible that the sector be a public utility with monopoly characteristics. In such a situation, the firm is the market and as long as the activity is specific to the sector, the monopoly may have the ability to internalize all the economies of specialization.

However, it is Williamson (1971) that had already stressed some of the obvious problems with the options considered by Demsetz, options such as the contracting problems that would arise if one of the competing firm were to monopolize through economies of scale one of the layers of the production chain. Such an action would create problems for firms who would need that output to compete downstream. It is implicit to Stigler's analysis that this is why, in many cases, some of the options Demsetz highlights are highly unrealistic.

It may be possible that the division of labor has at best a very marginal impact some sectors and that scale economies are almost insignificant. Baumol (1965, 1971) illustrates that possibility in the context of the live entertainment sector just as Adam Smith had already illustrated it using the agricultural sector as example. However, even there, say, in the entertainment sector, technological improvements in transportation has had some mitigating effects. In addition, technology has created through digitalization partial substitutes with almost zero marginal cost, substitutes that can be commercialized to the general public through CDs, through Internet services such as Apple's iTunes, and through peer-to-peer services such as www.musicdish.net. Similarly, Cheung (1973) illustrates the same division of labor using the production of honey in the United States.

While it may be that the division of labor is very marginal in some sectors – we gave two examples above – it is not null. In addition, if we consider the sectors a little bit more broadly, say, in the entertainment sector, the digital media, we find that the division of labor is very much a reality. This means that Demsetz's contention about the exhaustion of economies of scale leads him to reject the division of labor just like Williamson and, implicitly, Perry (1989). This is evidently inconsistent with Adam Smith's analysis and unacceptable for Stigler.

2.5.3. Perry

Perry (1989) interprets Stigler's analysis as a model to explain industry growth's impact on the sector's equilibrium, namely, how the division of labor combined with growth results in vertical disintegration. Perry uses a two-layer formulation where the upstream layer becomes a monopoly while the downstream layer disintegrates into a number of competitive firms. The only question left, he contends, is whether growth would eventually make entry into the upstream monopoly stage a possibility, suggesting implicitly that it might not, hence, our contention that he implicitly recognizes that the scope of the division of labor may be limited.

Perry is also skeptical about Stigler's unique focus on the scale dimension, noting that economies of scope may overwhelm the economies of scale. Perry's monopoly may very well be a multiproduct firm, a possibility that creates interesting questions where product differentiation tend to blur the sectors' various boundaries. This is a possibility that we will consider briefly below.

Finally, Perry contrasts Stigler's model with what he considers a far more conventional model. In the latter model, firms seek to expand and achieve benefits from economies of scale (Harrigan 1984). Their objective is not solely to achieve those benefits but as much to then expend in closely related stages with a view to build economies of scale and scope in those stages too. He recognizes that this scenario is based on a number of strong assumptions. For instance, those stages' economies of scale must not be significant until production achieve some sufficient level of production and it is only at that stage that they kick in. Even that assumption is insufficient and, as he notes, the separability between the two layers must be incomplete so that the firm that integrates those two stages would be able to extract the benefits from vertical economies of scope between those two stages.

The alternative vertical integration scenario Perry presents is based on questionable assumptions. For instance, it does not explain why the initial situation was a disintegrated environment. However, the same outcome could emerge in a different manner. For instance, there may be enough market imperfections, at various levels, say, at the financial level, and there may be enough risks that various entrepreneurs may assess the probability of various outcomes differently. If that is the case, then it may well possible for one entrepreneur to make better decisions – which one, we cannot know ex ante – and it may well be that that decision justifies a successful vertical integration across two or more layers the way Perry describes it. In that situation, the vertical integration is unlikely to be buttressed upon a greater efficiency and the entrepreneur's motivation is more likely to be the setting of entry barriers by increasing the entry cost for would be competitors. Such a strategy could be further reinforced if the same entrepreneur were to create additional entry barriers, say, through putting in place and managing asset specificities, using the market power it has gained.

The last scenario is of some relevance to Stigler's model because it can show how the original firm may be able to preempt, in some cases the vertical disintegration that follows from Adam Smith's proposition. In that case, we would have a scenario similar to the one describe by Demsetz. Demsetz does not explain why the various scenarios he describes would emerge. In many cases, those could not be based on transaction costs that are too high. They could be based, on the other hand, on the kind of scenario we have just outlined representing the effort by the original firm to protect its original position in spite of growth in the extent of the market. The possibility outlined here could be illustrated by the effort of various utilities, including local telephone companies, to keep the vertical control over facilities such as trenches, poles, conduits, and other real estate in spite of the greater efficiency that could be gained through a divestiture of those facilities.

2.5.4. Becker and Murphy

Becker and Murphy (1992) provide an interesting and innovative criticism of Adam Smith's proposition. Their objective is to show that the extent of the market is not the sole factor that limits the division of labor and that those other factors limit the division of labor probably far more often than the extent of the market.

Other than Yang, Becker and Murphy are the only one to formally develop a division of labor-based production model. The model they build combines features of the new classical approach with the neoclassical production theory. It has a number of interesting features that facilitate the analysis of a number of cost contributing elements that are only indirectly related to the production of the output. In many ways, one can identify a number of common points with Young's (1928) "indirect or roundabout methods of production."

There is a major difference between Young's and Becker and Murphy's treatment of those

costs, namely, the cost elements they choose to focus on. Becker and Murphy's objective is to show that constraints other than the extent of the market normally actually restrict the division of labor. They study more specifically two dimensions that are fundamental, they believe. The first relates to the costs one would expect from an increased division of labor. For them, the increased division of labor, since it is nothing less than a greater differentiation in tasks, implies inevitably a higher complexity level in management, i.e., coordination costs. The second relates to the costs associated with the growth in knowledge that coexists with the division of labor and the constraint it imposes on individuals who have to be able to process it if they are to be able to contribute positively to the division of labor process. It is interesting to note that the problems they choose to address are not essentially different from some of the questions Williamson had raised. Becker and Murphy, just like Williamson and Demsetz, set themselves to establish that the division of labor is interesting but that it is only one among a long list of factors that effectively restrict the extent of the division of labor.

Becker and Murphy use their division of labor production model to analyze more formally such factors as the workers' productivity, the level of knowledge, and the coordination costs traceable to extending the division of labor. They also take into account the learning constraints it imposes on workers who must allocate their time between education and work. The growth of knowledge combined with the division of labor imposes complementary educational constraints that requires that sufficient, ideally optimal time be allocated to training and education instead of direct production. It also means that some workers have to leave the production process to enter into teaching in order to transfer the expanding knowledge to the other workers so that they can become productive in the new environment characterized by a greater level of differentiation in their tasks at work. While their analysis is more formal, the approach is not different from Young's once one understands that training is nothing more than an intermediate product that reflects the response to the division of labor, one of these indirect effects Young discusses. The selection of those particular factors that are required to achieve that roundabout production process is what leads them to a conclusion that looks on the surface different from Young's and Adam Smith's.

However, there are significant limitations to their approach that relegate the scope and the importance of their conclusions to nothing more than what Williamson and Demsetz had already shown informally. We can illustrate their problem with the help of one example from their paper. They argue that Adam Smith's division of labor model would bring about, at its limit, a state where each individual within a market would have a different specialization from every other individual,i.e., that Adam Smith's proposition would require that we observe at least as many specialized trades as there are people in a given market. This is a natural outcome of the way they choose to specify Leontief's fixed proportion model.

In particular, were the division of labor actually limited by the extent of the market would mean, they argue, that, were we to consider pediatricians, each pediatrician would develop in a given market, say, Istambul, a distinct specialty in the field of pediatrics. Since this is not the case, they argue that this is due to the high coordination cost this would impose on the services pediatricians offer. The problem is different. It is akin to the problem we discussed earlier about the division of labor's technological limitations in relation to trades such as artists. The problem with pediatrics is somewhat of the same nature. As long as every parent train as pediatrician and provided they have a sufficient number of children to gain a minimum of the real life experience physicians need to practice, there is a need for a pediatrician with a general training to effect, at a minimum, the required triage among potential patients. Triage of anyone individual takes a minimum amount of time. In addition, as a pediatrician carries out the triage function, the time required for triage combined with required knowledge means naturally that, even with zero coordination cost, the division of labor would not mean what Becker and Murphy argue it means.

The argument above demonstrates also that their model is making them fall in the very trap Stigler (1951) fell in. They argue that the division of labor combined with the extent of the market can be usefully and, probably, better interpreted in terms of coordination costs. In this context, they argue that Adam Smith's conclusion that the level of specialization in the pin

factory was limited by the extent of the market has a problem. They argue that, had it not been because of coordination costs, the various pin factories Smith describes would have had an incentive to combine and create a monopoly "to get a larger scale and market, and specialize more within each factory." They ignore the technological characteristics of the processes that are subject to the division of labor and fall in the trap of equating economies of scale with the division of labor's increasing returns..

As Adam Smith noted, the division of labor and the increasingly narrowing of the specification of the tasks may often lead to the development of new machines, hence, new technologies that are inconsistent with their model. This is especially true where disruptive innovations is the solution that emerge with the expansion of the market. The process through which the division of labor works itself across sectors has been studied by Young (1928) while Robertson and Alston (1992) have analyzed in details how the division of labor would typically work itself through within a firm.

Just as Becker and Murphy argue that it is more appropriate to reverse the causality, i.e., to go from the coordination costs to limitations on the division of labor than to accept Adam Smith's causal analysis, they argue also that it is more appropriate to see the causality going from knowledge and investment in knowledge to limitations on the division of labor.

Becker and Murphy artificially set the problem to support their result by restricting the division of labor exclusively to labor specialization. Yet, Adam Smith had already considered the division-of-labor-based creation of new machines that are then integrated with the newly trained labor in the production process. That factor alone undermines their conclusion.

However, the most significant problem with their analysis is its static characteristic. It is intuitive that factors such as coordination costs and knowledge must influence the way the division of labor transforms the production process. One can imagine that it may take longer for the division of labor to follow its course. Through time, people learn to deal with coordination costs in a number of new ways including the division of labor combined with the codification of knowledge (Cohendet and Steinmueller).

2.6. Stigler's life cycle model

Stigler's analysis at the stage where he specifies the process through which innovative firms enter the economy, implement their innovation by successfully introducing new processes and producing new goods while learning in part to better adjust to some of the established processes in the economy. The process through which these new firms become established and their vertical structures take shape through vertical disintegration has everything to do with the division of labor. At the same time, the process that Stigler describes is not documented. One has no way to know what set of conditions and what set of incentives Stigler has in mind to establish how the innovative process will evolve into an industry and how a number of the activities that are carried out by those firms will eventually be taken by a new firm that will evolve through time into a new industry. This last step does not explain how and why the original industry will consider that it is better off by having that activity taken over by the market to enable it to focus on the kernel of the industry's original innovation.

We have seen that the first element of Stigler's approach to the vertical disintegration problem is based on overlaying the division of labor over some concept of the firm. If we complement Stigler's insight with Chandler's insight, the original industry would have an incentive to release certain activities to new, emerging sectors because, as Stigler describes it, the original firm, and, in time, the original industry carried out those activities to compensate for the absence of a market. It seems reasonable, again, using the insight of the process suggested by Chandler's analysis, that those were processes that the original sector did not have a comparative advantage in but that it had to pursue because, "[a]t a given time, these functions may be too small to support a specialized firm or firms. The sales of the product may be too small to support a specialized merchant;… the demand for market information may be too small to support a trade journal. The firm must then perform these functions itself" (p. 188).

Stigler is also implicitly assuming that the nature of innovations would be such as not to provide the innovator such market power as, say, to make the technology decisions sufficiently endogenous to support some market-power-enhancing path dependence. Such path dependence in telecommunications would seem to continue to be, at least at the local level, a major factor that is undermining the ability of entrepreneurs to build a competitive local infrastructure. Such path dependency can only have favored the ability of incumbents to increase the entry cost to new entrants while sustaining or even increasing such technologically sub-optimal, path dependence-based market power as seems to have been the case in telecommunications (Fransman 2002).

Stigler had to deal with the nature of the division of labor and its inherently dynamic, evolutionary nature that is foreign to the concept of a competitive general equilibrium.. He also had to buttress competitive dimension of the innovative process he was inheriting from the division of labor. Finally, he had to deal with vertical integration. He achieved all those objectives in an artificial manner by overlaying a Marshallian-inspired life cycle model.

Stigler's life cycle model consists of two transition phases between three stages. The stages are (1) the birth of the firm and the industry, (2) its maturity, and (3) its contraction while aging. The two phases are, first, while the firm is expanding, the process of disintegration and, as the industry contracts, the phase of reintegration. The rational for disintegration is well developed, as we have discussed it above. The rational for reintegration is even more contrived than Marshall's contraction phase. It is there to complete a picture and to provide the possibility that firm might integrate within a competitive environment but it is not supported by any serious analysis.

3. Division of labor

3.1. Introduction

The division of labor is central to the analysis of vertical integration. For instance, we will see that it implies a more sophisticated look at the production process than the neoclassical production/cost function. The production/cost function is typically specified ex post, i.e., it is irrelevant to the modeling of stakeholders' decisions. Where it is defined ex ante, on the basis of some technology set, it does not provide any information on how that technology set emerge and how it might evolve (Panzar 1989).

In addition, the division of labor is at the heart of the fault that necessarily exists between the classical and neoclassical approaches. The neoclassical model is inherently static and potentially embedded within a general equilibrium model. The comparative static analysis that is associated with it may be able to explain the equilibrium at each of the two reference points but it cannot explain what happens in the transition between the two equilibriums. The division of labor is by necessity a dynamic model. Young's (1928) brilliant insight leads him to position the dynamic movement between an equilibrium model à la Schumpeter and an evolutionary model à la Nelson and Winter. Young talks of "the counter forces which are continually defeating the forces that make for equilibrium are more pervasive and more deeply rooted in the constitution of the modern economic system than we commonly realize" (p. 533). The forces that make for an equilibrium are neoclassical, Walrasian, and/or Schumpeterian. The countervailing forces are evolutionary forces. In an earlier writing, Young had strongly rejected evolutionary analysis considering that only an equilibrium-based analysis could be justified. In 1928, he made a brilliant turn around, presenting the two sets of forces, those that push prices toward a long-run equilibrium, and those that emerge from pervasive external economies, as countervailing forces. Here he seems to suggest that he expects that, in general, the latter will dominate.

The division of labor has generally been conceptualized around and applied to the industrial sector. This is the way Xenophon, more recently, Adam Smith, and, still more recently, Young (1928) have primarily used it. In that context, the division of labor is viewed as the source of increasing returns. In other sectors, sectors such as agricultural and the service sector, the land constraint in the first case and the labor constraint in the other case can restrict the impact of the division of labor.

Adam Smith recognized agriculture's specificity, arguing that it was unlikely, for physical reasons to be able to support the same level of division of labor. Ricardo focused on the limited supply of land. He also realized that farmers would first occupy those areas where agricultural and transport conditions were the best to move through expansion to land that would progressively be of lower quality. This led him to conclude that the agricultural sector would be characterized by decreasing returns.

Until recently, the service sector had not been subjected to division of labor-type analyses. Nevertheless, the service sector can be expected to lead to interesting studies. Baumol (1971) made a comparative study of the economics of Athenian and modern theater production. In it, he showed how the increase division of labor at the level of actors reverses the cost benefits usually associated with it. More generally, Baumol's (1965) work on the art can easily and usefully be rephrased in the context of the division of labor. From the division of labor perspective it shows that the extent of the market needs not limit it. What Baumol's analysis points to is that wherever there is a service that is produced that cannot be disembodied from the individual who produces it, the individual artist, more generally, professional, may limit the division of labor long before the extent of the market could become a concern. Baumol did not position his work relative to the division of labor and yet his work is far more powerful and convincing about the exact meaning of Adam Smith's proposition.

One sector of the economy, the economy's infrastructure, has not been analyzed with the help

of the division of labor. It might sound surprising given its significance in Adam Smith and Mill's works. For Adam Smith, economies that had efficient governments – he had in mind countries such as the Netherlands - and that were provided with favorable transportation conditions – such as England and the Netherlands - were likely to extract greater benefits from the division of labor than other regions. Adam Smith's analysis led him to see the government as an agent that should be proactive in the provision of services such as education and national defense. The role of national defense was to protect trade, especially foreign trade, not to favor British merchants. It should also be used to open new regions to trade especially where governments were hostile to merchants and to trade. Mill was the first one to realize the unique nature of public utilities.

Young in Schumpeter et al. (1914) understood clearly the role of infrastructures and their contribution to the economy. He understood also the danger of underinvestment under an unregulated market system. He criticized proposals where railroads would be entitled to charge "what the traffic will bear," noting that such an approach "assumes a static view of the facts; it postulates that a large part of our railway plant and equipment is a given quantum, more than ample for present transportation needs" (Schumpeter et al. 1914, p. 84). He went on to illustrate the problem with the help of examples that all took a dynamic view of the economy and of the interaction across sectors. XXX has provided a unique historical perspective utilities can create (2000).

The effective constraint is a demand constraint reflected by the extent of the market. Labor is divided into more and more specialized functions and, through that process, it becomes increasingly efficient and, as a result, it succeeds in meeting increasing demand levels. Whatever the level of output, entrepreneurs will be able to identify new specialties for workers and new organizations to coordinate the work of these new specialized workers in such a way that, proportionately, the output increases more rapidly than the inputs.

The division of labor means that, for some output level, there is a level of specialization that is optimal, defined in a manner somewhat comparable to Stigler (1951) (Robertson and Alston 1992). It means that, as the output level changes, the firm will, at some stage reorganize while creating new specialization.

3.2. Increasing returns and externalities

3.2.1. Introduction

In what is probably the most humorous section of the paper, Stigler told us in 1951 that, "In 1928,… the neglect of increasing returns had gone so far that Allyn Young felt the need to restore perspective by an emphatic indorsement of the fundamental importance of Smith's proposition…" adding that Young's "position seemed persuasive, but he did not resolve the technical difficulties of incorporating the extent of the market into competitive price theory" (p. 187). Stigler's statement is surprising in two ways. First, the problem of increasing returns to scale had been the center of one of the main controversies of the 1920s. Then, Young, in the first paragraph of his paper, was of the opinion that neoclassical analytical tools, hence those tools that were to become the cornerstone of Stigler's price theory "may stand in the way of a clear view of the more general or elementary aspects of the phenomena of increasing returns" (p. 527).

Increasing returns to scale problem emerged as a central issue in Marshall's partial equilibrium analysis of competition. Marshall had to confront a number of problems to be able to construct his partial equilibrium analytical approach. The most serious among those was the problem of economies of scale. As we saw earlier with Stigler, Knight, and Sraffa, economies of scale in a sector imply a monopoly. Yet, Marshall, just like Stigler in 1951, was not satisfied by either ignoring those economies of scale or working on the basis of an economy with a predominance of monopolies. Rather than following Cournot's approach based on accepting some form of imperfect competition, Marshall chose a different approach, approach that has been mostly misunderstood by economists since then (Newman 1960). He suggested that economies of scale

could be subdivided in two different forms of economies, depending upon whether they were internal to the firm or external. Internal economies of scale are those economies that are fully reflected in the firm's decision process.

Marshall's external economies are economies that have not been internalized by any firm, hence, properly integrated in their decision process. They benefit the industry as a whole, Marshall paid particular attention to the localization of firms, greater concentrations providing benefits to all the firms in that locality without affecting the technological characteristics of the individual firms.

Marshall's external economies of scale resulted in a polarization among economists that is still surviving. On one side, following Pigou, and then Sraffa, the neoclassical model was increasingly formalized into Knight's theoretical formalization of neoclassical economics to become, then, internalized into the Chicago School. On the other side, following Young, we have the modern endogenous growth theory with Romer (1986) and others.

The fault that separates the two sides has its roots in the divide between static and dynamic analysis. Even though two of the key, early protagonists were Young and his student, Knight, it is as if each side were speaking a different language. If it were for no other reasons, Stigler's 1951 paper would still be important for building a bridge to help better understand the issues. In light of this fault and the lack of methodological steps they could share to build a bridge between their respective approaches, we have chosen to deal with the two approaches separately, starting with the neoclassical approach to then present the dynamic model introduced by Young (1928).

3.2.2. Neoclassical approach to external economies

The cornerstone of neoclassical analysis is allocation efficiency, i.e., insuring that agents receive the proper information through market signals to choose how to allocate their resources. The framework is essentially static. How the role of static comes to play in the scale economies is well illustrated by Knight's (1928) assessment of the neoclassical competition market failure.

Knight, just as Sraffa, and Pigou (1912) before them chose in Marshall what they wanted and ignored the rest, i.e., what could not be reduced to their simplified neoclassical theory. Marshall's external economies of scale could not elegantly be fitted in that model. Neoclassical analysis would have needed an industry-level partial equilibrium where the summation of the parts (i.e., the firms) would have been inadequate. That led Knight (1925) to set the following challenge for whomever would defend Marshall's external economies: "To vindicate decreasing costs ['*without tending toward monopoly*' (from p. 331)]… it must be shown that there are, or may be, industries, in a condition of stable competition, in which no producer already engaged could decrease his real costs by expanding his output at the expense of other producers, and yet in which real costs would be decreased all around by new producers entering the industry in competition with those already there." Noting that those new producers would be bidding up prices, hence, costs, he adds, "These inevitable sources of increasing cost must be more than offset by some kind of purely 'external' economies in organization." Knight (1925) concluded that, "I have never succeeded in picturing them in mind, or finding any convincing reason to believe they exist." (p. 332).

Knight (1924) did away with Marshall's externalities without seeking some other intermediate solution, say, along Cournot's imperfect competition and Sraffa (1926) buried it by offering a clear, bipolar approach. The need for formalization led to an analysis that incorporated a theory of perfect competition and a theory of the monopoly. Knight was pursuing the same path, positioning the neoclassical theory model as the cornerstone of economics. Thus he consistently argued that "the general theory of a subject is not supposed to give a realistically complete account of it. The function of theory is to identify, isolate, and trace out to their final consequences the more fundamental tendencies discernible in a complex mass of phenomena" (Fisher, Knight, and Parry 1921, p. 144). Knight did not question that neoclassical

analysis was what constituted the proper "general theory of economics" as seen, for instance in a further comment in 1921: "And the theory also shows the manner in which any society, even an ethical Utopia, would have to impute income to productive factors in order to organize production intelligently" (Fisher, Knight, and Parry 1921, p. 145).

Recently, Liebowitz and Margolis (1995) concluded "that the empirical importance of network externalities, as externalities, has been greatly overstated." While minimizing the impact of externalities, very much in the Knightian tradition, they acknowledge that "some networks, such as the network of English speakers, seemingly cannot be owned" (p. 7) yet, there is a considerable empirical literature that highlights the significance of languages on international trade (Head and Ries 1998). For instance, Rauch and Trindade (2002) suggest "that business and social networks have a considerable quantitative impact on international trade." Evidently, they add that the impact is achieved through "helping to match buyers and sellers…, in addition to their effect through enforcement of community sanctions that deter opportunistic behavior." Clearly, in most cases, a common language lowers the transaction cost just as the cultural ties among the members of the community. But, as Liebowitz noted, those are economies that are not internalized. This is but a small example of all the externalities that exists and that are rarely internalized. Where it is internalized, Shaw (1916) gave us, with Pygmalion, a taste of the kinds of transaction cost it involved even under the most ideal conditions.

Knight's criticisms were parts of the debate that followed Pigou's 1912 *Wealth and Welfare*. Marshall's concepts were not as formalized as Pigou's in part because Marshall had some misgivings about what might be lost in such formalization (Bharadwaj 1972). Marshall saw limitations in the static analysis of increasing returns and saw those limitations in Pigou's work.

The challenge of Marshall's external economies is part of a much more serious chasm between classical and neoclassical analyses. Knight's position is that constant returns can be expected to be the norm rather than the exception.

3.2.3. Classical analysis and external economies of scale

Young's (1913) review of Pigou's 1912 *Wealth and Welfare* is well known because it started a lively literature on the internalization of externalities that arose in the context of Marshall-like externalities. Knight, his student, pushed the attack much further challenging the very concept of Marshallian external economies. While Young and Knight both agreed on Pigou's failure to take into account the internalization of the Marshallian externalities, it is important to note the difference of perspective between the two of them. We saw earlier that Knight took constant returns to scale as the norm in economic analysis, a hypothesis that simplified significantly his adoption of neoclassical analysis. On the other hand, Young (1913) states: "in cases of diminishing and of increasing returns, that is to say, in the generality of cases…" (p. 680). Where Knight believed in equilibrium and stability and permanence, Young acknowledged the market forces Knight was focusing on while pointing to the countervailing forces of the division of labor. Knight's world appears devoid of innovation. That difference is at the center of this paper. As the Chicago School was largely the child of Knight and as Stigler was one of his best known student, Stigler never succeeded to sufficiently deviate from the Knightian framework to tackle Young's contribution. In spite of the Chicago stranglehold, Stigler had enough intuition to understand that Smith's proposition combined with the analysis of Marshall and Young could help him where neoclassical analysis was powerless.

Prendergast's (1993) view is that the classical economists, not just Smith but, already before him, Petty, Martyn, and Mandeville, saw increasing returns as a function of the general economic progress much more than as a function of scale. He points out that competition pushed firms to innovate through the division of labor in order to keep pace with its competitors. Marshall justified his external economies (i.e., those economies "which do not directly depend on the size of individual houses of business" Marshall as cited by Prendergast) largely upon geographical consideration, considering the role of greater concentrations of some trade. While we have identified other factors that can contribute to external economies, the

localization of industries is still a major factor in explaining large external economies that are at best only partially internalized. Examples of studies that confirm Marshall's broad hypotheses, even if they disagree with one another include Kenney and von Burg's (1999) and Saxenian's (1999) analyses of Silicon Valley and Route 128.

…

3.3. Increasing returns and economies of scale

Stigler starts his paper by explaining how Adam Smith's proposition would result in a monopoly in every sector. At this initial stage, he develops his monopoly argument by considering what would happen if the particular sector was made up of a number of firms. It is reasonable to infer that he has in mind a situation such as Smith's pin factory. We can imagine some markets, say, Glasgow and Edinburgh respectively. Each would be served by pin factories, let's assume only one in each city. We can now consider some sufficiently improved transportation between those two cities that is sufficiently significant that those two cities become now a single market for pins. At this stage, Stigler argues that one of these two factories can be expected to be more responsive to the change in the extent of the market and expand its activities to sell pins not only in its own city but also in the other city. That firm will effectively have a larger output. This means, if we follow Stigler's argument, that this same firm will then be able to benefit from the division of labor by increasing the number of stages in its process while achieving greater specialization. Stigler tells us that that firm will be able to lower its cost as a result of the greater specialization. Its cost advantage enables it to compete with the other pin factory and expand further its output. It also means that the other firm may have to forego some level of specialization. Under those conditions, we are eventually left with a single firm, i.e., Stigler's monopoly.

The division of labor implies increasing returns to the extent that the only incentive there is to specialize further is to lower cost per unit of output. It follows that, from that perspective, the division of labor's increasing returns look very much like economies of scale, observation Stigler made in 1976.

The analysis of increasing returns we just presented is inconsistent with Young's (1928) analysis:

> "[T]he principal economies which manifest themselves in increasing returns are the economies of capitalistic or roundabout methods of production. These economies, again, are largely identical with the economies of the division of labour in its most important modern forms. In fact, these economies lie under our eyes, but we may miss them if we try to make of *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 firm… the economies of roundabout methods… depend upon the extent of the market – and that, of course, is why we discuss them under the head of increasing returns. It would hardly be necessary to stress this point, if it were not that the economies of large-scale operations and of 'mass-production' are often referred to as though they could be had for the taking, by means of a 'rational' reorganization of industry" (p. 531).

In the above quote, Young clearly argues that the increasing returns the division of labor creates may on occasion be associated with economies of scale but that this is only a secondary effect. The increasing returns, Young argues, are effectively uncorrelated to the size of the firm. Earlier he had chided "economists of standing" for having "suggested that increasing returns… where they are present they must lead to monopoly" (p. 531).

Effectively, Young is challenging the whole foundation of the technologically and organizationally determined monopoly as an inevitable reflection of increasing returns-based greater efficiency. In other words, Young is challenging the direct link between increasing returns and economies of scale.

At this stage, it is important to note that Stigler had also challenged the conventional Sraffian analysis but to a much smaller extent. Stigler's 1951 result means that it is not because a firm is characterized by what looks like economies of scale that one can assert that the firm's technology and organization are the source of ex ante economies of scale. Stigler's point was that those economies could be only present in a subset of the firm's activities. It was that, absent a disaggregated analysis of the firm's various activities, one could only identify ex post economies of scale. For Stigler, the proper level to discuss economies of scale, hence, the monopoly, is not the firm but the individual activity.

Young's attack is far more fundamental. To understand the basis for Young's attack, one needs to go back to the division of labor. The division of labor is a two-step procedure. In the first step, the division of labor replaces one production process by another one that is characterized by greater specialization. In the second step, the transformed, newly specialized inputs, possibly labor, possibly machines (ideally seen from the kind of perspective Babbage had already introduced, i.e., systems that integrated processes and machines and labor into a coherent, efficient organization - Schaffer 1994), are then inputs into the new production process.

The division of labor stresses the increasing returns that is gained from the transformation of the inputs. It conveys little information on the second stage, i.e., the effective production function. The point Young is making is that the latter need not be associated with either economies of scale or large firms.

Young's argument stresses the problem with the neoclassical production (cost) function that is effectively an ex post description of the organization and of the technology. From Young's perspective, it is an aggregate that is made up of two totally distinct elements, first the transformation of the inputs through the division of labor and then the transformation of the way those inputs are combined to produce the output. Young's analysis stresses that the division of labor cannot be reduced to fit the neoclassical analytical framework without being eviscerated of all meaning. Referring to the earlier quote, p. 531, Young had established in 1928 that "increasing returns… where they are present" need not and, generally, will not "lead to monopoly" in spite of Knight's (1925), Sraffa's (1926), Stigler's (1951) and everyone else assertions to the contrary.

4. Vertical integration: "The division of labor is limited by the extent of the market"

4.1. Introduction

4.2. Competition

Young (1928) and Stigler (1951) both assume that competition prevails. This is the standard hypothesis the economic literature makes when considering vertical integration (Coase 1937; Williamson 1975). This is the only way to carry the analysis without concerns that firm-level anticompetitive strategies may distort the results. Coase has been more explicit about this hypothesis than others. For instance, in "The Nature of the Firm: Origin" in 1991, he specifies: "I mean apart from attempts to obtain monopoly," (Coase 1991, p. 40). In 1937, he had far more specific, imposing a strong condition: "The entrepreneur has to carry out his function at less cost, taking into account the fact that he may get factors of production at a lower price than the market transactions he supercedes, because it is always possible to revert to the open market if he fails to do this" (Coase 1991, p. 20). The latter condition, namely the option to return to the market, implies that the firm is small enough not to affect the market's efficiency. It also means that it does not apply in a sector such as telecommunications where the markets have not existed over a very long time, i.e., where whatever benchmark might have existed does not exist anymore.

There is another way to specify Coase's condition, namely to impose the common neoclassical hypotheses of perfect information and zero transaction costs. However, Coase's hypothesis is insufficient. The firm that integrates vertically must not be able to impact the overall technology or, if it does through some innovation, then it must not be able to control the technological path to her strategic benefit. It is easy to understand such an additional assumption by looking at the telecommunications sector from a more historical perspective. Through their monopoly, whether conscious or not, incumbents have been de facto imposing their preferred technological and organizational structure upon the industry. This means that incumbents have been imposing a certain path dependence on the technology of the sector, the sector's organization, and the firm level transaction costs. It also means that asset specificities and other ways to protect against opportunism are, a priori, the result of decisions made by incumbents. This means that the causality path that traditionally takes those factors as exogenous must now look at them as variables incumbents are able to manipulate to further their own private goals. Those conditions are essentially different from those that apply in the vertical integration literature, say, those found in Williamson. The literature is consistent with Stigler's *Price Theory* framework but not with the 1951 paper.

Stigler (1951) tells us essentially nothing about what he means by competition. However, as his frame of reference is Smith, it seems reasonable to start from Smithian competition. For Smith, competition cannot be treated in isolation of the division of labor. At the same time, if we use modern terminology, the closest synonym to the division of labor would seem to be innovation. The division of labor represents factors such as a reorganization of the tasks, new tools, and new working conditions, all of whom correspond to achieving an improved process (and potentially a new output), all elements that go to defining innovation. Firms stay ahead of their competitors or catch up after them by finding new ways to implement and take advantage of the division of labor, say, though such moves as vertical disintegration.

Thanks to Rosenberg (1976), it is possible to offer a far more specific definition of Smithian competition. The source Rosenberg uses is not the Wealth of Nations. Rather, it comes from a series of lectures Smith delivered in 1762-1763 at the University of Glasgow and who were recently published. Rosenberg's analysis does not come from *The Wealth of Nations*, yet it complements the analysis carried out by Smith in *The Wealth of Nations* and make the book

that much more interesting and path-breaking. The subject we can consider here is the one developed by Rosenberg, namely, how making it possible to apply the division of labor to judges led to a significant improvement in justice. The division of labor applied to the judicial system led to replacing, say, Roman generals who, Smith tells us, cumulated also the position of judges. Smith adds that their position as general was what gave them fame and wealth and recognition. The position of judges was secondary to the system whereas, later as in the time of Smith it had evolved into a full-time profession.

Quoting Rosenberg's (1976, p. 866) discussion of Smith's notes: "[w]hen an individual performs more than one function, but his success is based upon only his performance in one of them, he will neglect the one to which his social status is less attached. On the other hand, it is important that social functions not be reposed in large social aggregates. A division of labor involving fewness is essential because fewness is a precondition for establishing individual responsibility. Only when individual responsibility (or fault) can be established will individual effort be reliably engaged in ways which will promote society's larger interests."

While Rosenberg's analysis deals only with the judicial process, it can be extrapolated to competition, hence our reference to "Smithian competition." Smith's key argument is that judges must be accountable to society for their decision. Accountability was essential, in Smith's framework, to insure that the judges would inform themselves properly of the specific of the cases they had to judge, that they put enough dedication to their job, i.e., that for the judicial system to function efficiently. To repeat part of the previous quote, "Only when individual responsibility (or fault) can be established will individual effort be reliably engaged in ways which will promote society's larger interests." Rosenberg continues to stress the point observing that the division of labor meant that judges would focus on their primary responsibility, providing fair and predictable decisions. Then, if judges were named to serve a large region, this would mean that people in society would not know the judges and would not be able to develop a stable benchmark for justice, a requirement to trust the system.

Fewness could not mean too small number. Fewness must facilitate the provision of a benchmark, even if it is informal. In terms of competition, fewness means that buyers are able to gather information about sellers that may impact their choice. The information could relate to characteristics such as the quality of the product sold, the quality of follow-up service, and some reasonable of product differentiation. In the time of Smith, fewness meant local judges. In today's commerce, fewness may be related to branding. Once a product is successfully branded, the brand would suffer from providing poor products and poor services. At the same time the number of brands must be small since people's capacity to remember a large number of names is limited. Branding might meet Smith's criteria in as much as there is a sufficient number of brands to make it almost costless for individuals to shift from a brand, had they received poor service.

It is easy to see that Smith was providing a very powerful competitive framework that has wide applicability in today's economy. In making this statement, we are suggesting that commercial branding may be acting as a substitute for local providers. We can illustrate how Smithian may be working through the example of retail channels. Retail channels appear to be an industry structure that is characterized by large scale economies, yet, it is also an industry structure were the concentration level appears to be, from a heuristic perspective, consistent with Smithian competition. Now, if we look in wireline telephony and cable, we observe a vertical retail structure that would suggest a lack of competition.

The complementary dimension to the local criteria Smith had in mind, criteria we reinterpreted in a modern setting as branding is innovation. We know from the world of bazaar that the number of players and their local availability do not guarantee competition. Sellers are reluctant to undercut each other prices. The instrument Smith had in mind to expand accountability and to ensure that competition brought benefits to end-users is innovation. Innovation creates some level of uncertainty that forces stakeholders to be alert rather than accepting passively the status quo.

4.3. The Nature of the Firm

Given Smithian competition, we can now compare the Stiglerian firm with the Coasian firm. The Coasian firm treats the firm as a set of interdependent resources – one could imagine a financial investment portfolio. It tells us that, given a mix of activities that together form a firm and given existing markets, it is not efficient to add or subtract activities. It is better to complement internal In addition, all the activities are taken as standalone activities and do not take into consideration the entrepreneur and her unique ability to organize those activities.

The Coasian firm is, in the static sense, a Coasian firm. Coase tells us that, in 1932, "I illustrated my position with diagrams somewhat in the manner of Stigler in his 1951 article" (Coase p. 40). Where the Stiglerian firm is far more sophisticated and complex than the Coasian firm is in the dynamic dimension that Stigler introduces in his third model. We discover a firm that enters the market. We can presume that it is to implement an innovation, to find customers for whatever new product or service it is offering, and as it expands to attract other firms with technologies that may be somewhat different or firms that are innovating through imitation, say, reverse engineering.

…

4.4. Externalities

One of Stigler's contributions was to introduce the firm in the context of Adam Smith's proposition. Another one was to introduce it as a new entrant that enters because it is innovating. One of Young's (1928) contributions was to generalize Smith's division of labor around the transformation of the process' inputs. Another one was to build upon that generalization to sketch a division of labor-based general disequilibrium dynamic framework. Another one was to identify the Youngian entrepreneur's mercantilist incentive. Still another one was to identify forces such as his mercantilist entrepreneurs that would act as a multiplier pushing along and possibly further amplifying the external economies. In this process, he did not properly appreciate the value of a firm-level analysis in pushing his classical vision. As for us, by taking Stigler's contribution and resetting it in

By introducing the firm within Young's analysis, we are able to solve Stigler's problem and formulate an approach that could for the basis for a division-of-labor-based theory of the firm within the context. Young understood the difference between division of labor and production functions, increasing returns and economies of scale, dynamic and static analysis. Young's contribution is best understood through Zeno's paradox: the only proper description of the economic process is within its dynamic development. For Young, Sraffa's monopoly problem is just an artificial construct of neoclassical analysis without general applications. Classical analysis provided the proper framework and, within classical analysis, the monopoly problem does not have the link to increasing returns scale has to the classical framework.

The neoclassical monopoly problem is primarily the result of two factors, the static model of the framework and the ex post characteristic of the production function. Considering the second factor, the contribution of the division of labor is essentially different from the specification of the production function. The division of labor is a two-step process that involve in turn the increasingly specialization of the factor inputs. In the first step, one models the transformation of the factor inputs, not just labor but also capital through machinery, and organization. It is only in the second step that the production function is introduced. The economies of scale are known to exist in the specialization of the inputs. However, there are no a priori reasons that they be playing a role in the division of labor-based physical process.

The transformation of the inputs due to the division of labor introduces by construction economies of scale. In Xenophon's and Smith's examples, the same worker sees her productivity multiplied through increasing specialization. Those economies of scale are imposed by the technology and organizational characteristics of the division of labor and they

are not under the direct control of the firm, at least in sectors where the firm does not have the market power to control technology. On the other hand, it is those new, transformed factors of production that, in turn, are introduced in the (intermediate or final) commodity(ies). The training and learning-by-doing processes are input specific and the production function is based on already transformed inputs. This means that the sector can achieve substantial scale economies ex post without the production process having to have such characteristics.

Young's (1928) complementary contribution is in further developing Marshall's external economies of scale. Young observed that firm or even sector-specific division-of-labor-based innovations tend to have an impact that easily expands far beyond the firm or even the sector and that impact is largely not internalized and, typically, cannot be internalized. This can best be illustrated by the impact of the division of labor on the specialization of worker and on machinery. The specialization of labor for a firm or even in an industry informs other sectors of new possibilities for their own labor. In addition, the training of the labor that is required to achieve the benefits of the division of labor's efficiencies can never be fully internalized. In 1963, Rosenberg studied the emergence of the machine-tool sector. Rosenberg's results show that the sector has simplified through time as the industry emerged and grew. He justifies it by observing that there is a finite number of tasks machine-tools are used for and that those tasks are independent of the industry where they are used. Rosenberg's result illustrates well the kind of externalities Young had in mind in 1928.

5. A model of innovation-based vertical integration

We are now in a position to revisit Stigler's vertical integration and address the problems we have identified in this paper. Our approach consists in purging Stigler's model of its neoclassical elements, elements such as the firm's life cycle model and the concept of vertical integration, to reposition it in the context of Young's analysis. building upon a model of competition derived from Smith. The vertical integration framework we consider is bounded by the division of labor that is fully embedded within the process of innovation. In other work, we do not consider the generic, static and accounting-based portfolio model of vertical integration à la Coase.

We consider a firm that is innovating either to compete or to enter. The innovation could be situated at an intermediate or the final stage of production. We take as given, for simplicity, the reorganization process and the time it takes to be implemented and we consider the firm reorganized. Nevertheless, we focus on the firm's decision process, in this transition period, especially the firm's "buy vs. build" decision process. We recognize that the firm has an incentive to specify its complement strategy, including its "buy vs. build" decisions, with a view to use its innovation to gain a competitive advantage toward building market power. We assume that the firm's ability to be successful in this endeavor is minimal. This is illustrated by Edison's attempt to impose direct current as the standard. One of the justifications for our assumption is that innovation is a risky investment, i.e., that innovators have an incentive to minimize risk, i.e., to minimize capital expenditures in favor of expenses. For that same reason, innovators do not generally have an incentive to in crease their investment to be able to block competitors in the eventuality that they may be successful.

The firm is made up of a new set of activities, very much along the Stiglerian/Coasian model. To achieve that stage, the firm is expected to have to complement its reorganization with new inputs, i.e., new labor with new kinds of capabilities and new level of specialization, new types of material, and new machines, as well as new requirements on quality and supply conditions. One way to think about our setting is to think about the steel mill that was integrated within Ford's River Rouge plant when it was first open (Chandler; Hughes). We reject Knight's end-to-end approach, especially where it may involve infrastructures, especially infrastructure that are organized in such a way to ha cannot be efficiently provided within Smithian competition. Where we cannot neglect the infrastructure as, say, the government infrastructure, we assume that they are independent of the innovative process, i.e., that the innovator (or its competitors or upstream stakeholders including government policies) has no incentive to allocate resources to influence the infrastructure provider, including the government to create distortions that

may affect the innovation. The firm's innovation may be partially/completely protected by a set of patents. We assume that none of those patents are such as to prevent or delay significantly competitive entry through somewhat different technological path and/or through imitation. It follows that the innovation will affect the technological/organizational/ institutional path but that neither the innovators nor its competitors would be able to affect that path to their advantage.

These activity conditions enable us to adopt a modified version of Stigler's second model. The modifications include, first, that the activities are those that emerge from the reorganization, and, second, the division of labor. The latter means that we need to consider on the one hand the newly specialized labor in each activity as well as the new inputs at both the firm level and the activity level, in the form of material and machines, software, etc. as distinct from the old inputs. It is required at the activity level in order to be able to model the activity's production/ cost function and it is required at the firm level to differentiate between inputs and internally produced intermediate products.

Following Stigler, it is reasonable to consider that some of these activities could at some stage be performed by different firms - this was evident in the case of the River Rouge steel mill. In other cases, until an entrepreneur come with the idea, it may not be possible to know ex ante that the firm is vertically integrated. In general, vertical integration would seem to be a misnomer when applied in the context of innovation.

We can transform Stigler's life cycle model of the firm into a life cycle model for innovations. The life cycle model of innovation is inevitable in as much as we accept the essentially dynamic nature of the Smith-Young model, a model that produces innovation through the division of labor both through the extent of the market and through competition among firms. The innovation life cycle model summarizes in a simplified manner the idea that an innovation is new and different when it emerges but that it tends to age through time as other, newer innovation emerges. The innovation can be thought as maturing into what can be considered as the norm. Then, obsolescence takes over progressively.

Stigler's analysis of the growth of the new entrant, in his third model, need to be also amended at two levels. First, we note that innovations are not always independent of one another. Then, the innovation, in Stigler's model, that primes the disintegration cycle can be introduced by an established firm or by a new entrant. An established firm would have an incentive to use it as a necessary competitive response within the Smithian framework we are considering.

External economies, especially, Youngian external economies are not essential to prime the vertical integration process within a Smithian competitive framework. While they would help amplify it, innovations are sufficient to keep the process alive. Nevertheless, Youngian external economies further the dynamism of the Smithian competition. However, Youngian external economies may be essential to show that the Stiglerian-Smithian-Youngian vertical integration process needs not be exclusively a disintegration process as implied by Stigler.

Young (1928), in his discussion of increasing returns, notes that, "the enlarging of the market for any one commodity, produced under conditions of increasing returns, generally has the net effect, as I have tried to show, of enlarging the market for other commodities. The business man's mercantilistic emphasis upon markets may have a sounder basis than the economist who thinks mostly in terms of economic statics is prone to admit. How far "selling expenses," for example, are to be counted sheer economic waste depends upon their effects upon the aggregate product of industry" (p. 537, italics added). It is clear that the "mercantilistic" process in and of itself is internalized by the firm. At the same time, as the innovator reach a broader range of potential customers, the more likely it is for the mercantilistic effort to expand the range of externalities beyond those that are predictable, i.e., externalities such as the government and many of the infrastructures.

Having resituated Stigler's analysis within the Smith-Young economic model of the division of

labor, we are able to deal with the problems that undermined Stigler's solution. The Smith-Young model does away with the monopoly model that is inherent to neoclassical analysis. We shifted Stigler's second model from a neoclassical framework to a Young-Smith classical framework, based on the division of labor.

We have then revisited Stigler's dynamic model, focusing on the link between the division of labor and innovation. By transferring the life cycle model from the firm to the innovation, the analysis starts with a firm that is either established or entering and that is innovating to be able to remain competitive.

We follow Stigler's analysis of the innovating firm since his arguments are developed in terms of the division of labor-based dynamic model that reflects an innovation's life cycle from its inception to its maturity. As it matures, the innovation is more and more accepted, resulting in increased demand. An industry emerges through competition and some entrepreneur innovates and takes over one or more of the activities of the original firm. The process is the result of the division of labor

BIBLIOGRAPHY

Alchian, Armen A.and Demsetz, Harold. "Production, Information Costs and Economic Organization", *The American Economic Review*, 62 (5), 1972. 777-795.

Ames, Edward and Rosenberg, Nathan. "The progressive division and specialization of industries",

Arora, Ashish and Gambardella, Alfonso. "The changing technology of technical change: General and abstract knowledge and the division of innovative labor". *Research Policy*, 23, 1994. 523-532.

Arrow, Kenneth J. "The economic implications of learning by doing", *The Review of Economic Studies*, 29 (3), 1962, 155-173.

Arrow, Kenneth J. "Methodological individualism and social knowledge", *The American Economic Review*, 84 (2), 1994, 1-9.

Aslanbeigui, Nahid and Naples, Micheles I., "Scissors or Horizon: Neoclassical Debates about Returns to Scale, Costs, and Long-Run Supply, 1926-1942', *Southern Economic Journal*, 64 (2), 1997, 517-530.

Barber, William J. *A History of Economic Thought*. 1967, originally published by Praeger 1968 and Penguin 1967. This webpage, http://www.wesleyan.edu/css/readings/Barber/***toc***.htm, has been authorized by William J. Barber as of October 23 2002.

Barnett, George E. "Chapters on Machinery and Labor", *The Quarterly Journal of Economics*, 40 (1), 1925, 111-123.

Baumol, William J. and Bowen, W.G. "On the performing arts: the anatomy of their economic problems". *The American Economic Review*, 55 (1/2), 1965. 495-502.

Baumol, William J. "Economics of Athenian drama: its relevance for the arts in a small city today", *The Quarterly Journal of Economics*, 85 (3), 1971. 365-376.

Becker, Gary S. and Murphy, Kevin M. "The division of labor, coordination costs, and knowledge". *The Quarterly Journal of Economics*, 107 (4), 1992, 1137-1160.

Boothe, Jason A. "Historical antecedents of contemporary endogenous growth theory",

Breshnahan, Timothy and Gambardella, Alfonso "The division of inventive labor and the extent of the market", 30 May 1997.

Bresnahan, Timothy F. and Gambardella, Alfonso. "The division of inventive labor and the extent of the market", Working paper, Stanford, CA: Stanford University. 20 January 2005: (www-econ.stanford.edu/faculty/workp/swp97029.pdf).

Buchanan, James M. and Stubblebine, W.C. "Externality", *Economica*, 371, 1962,

Chen, Ping. "A Biological Perspective of Macro Dynamics and Division of Labor: Persistent Cycles, Disruptive Technology, and the Trade-off between Stability and Complexity", Working Paper, 18 September 2003. 20 January 2005: (www.vwl.tuwien.ac.at/hanappi/PChenBiopes903.pdf).

Chipman, J. 1970. "External Economies of Scale and Competitive Equilibrium." *Quarterly*

Journal of Economics, 134:347-385.

Christensen, Clayton M. *The innovator's dilemma*. New York, N.Y.: HarperBusiness, 2000.

Ciliberto, Federico and Panzar, John C. "The vertical organization of a competitive industry. Theory and evidence from 19th Century Lancashire". September 2001.

Cohn, Steve. "Telling other stories: Heterodox critiques of neoclassical micro principles texts". Global Development And Environment Institute (GDAE) Working Paper NO. 00-06, Tufts University, August 2000.

Dahlman, Carl J. "The problem of externality", *The Journal of Law and Economics*, 141-162.

Davis, Lewis S. "The Division of Labor, Coordination Costs and the Growth of Government", Working Paper 1998-03, Oswego, NY: SUNY Oswego, 1998.

Demsetz, Harold. *Ownership, Control, and the Firm: The Organization of Economic Activity*. Vol. 1. Oxford, U.K.: Basil Blackwell, 1988.

Demsetz, Harold. "George J. Stigler: mid-century neoclassicalist with a passion to quantify". *The Journal of Political Economy*, 101 (5), 1993. 793-808.

Diamantaras, Dimitrios and Gilles, Robert P. "On the Microeconomics of Specialization", Virginia Tech Working Paper E2002-3, April 2002.

Edwards, Brian K. and Starr, Ross M. "A note on indivisibilities, specialization, and economies of scale". *The American Economic Review*, 77 (1), 1987, 192-194.

Faunce, William A. "Automation and the division of labor", *Social Problems*, 13 (2), 1965, 149 -160.

Fligstein, Neil. *The Architecture of markets: An Economic Sociology of Twenty-First-Century Societies*. Princeton, N.J.: Princeton University Press. 2001.

Foss, Nicolai J. "The new growth theory: some intellectual growth accounting", DRUID Working Paper, 1996.

François, Joseph J. "Producer services, scale, and the division of labor", *Oxford Economic Papers*, New Series, 42 (4), 1990, 715-729.

Fransman, Martin. *Telecoms in the Internet Age: From Boom to Bust to …?* Oxford University Press. October 2002.

Garicano, Luis and Hubbard, Thomas N. "Hierarchies, specialization, and increasing returns to knowledge: theory and evidence from the legal services industry". March 2004 .

Gasmi, Farid, Kennet, David Mark, Laffont, Jean-Jacques, Sharkey, William W.2002, *Cost Proxy Models and Telecommunications Policy*, MIT Press: Cambridge, MA. ISBN 0-262-7237-8.

Gibbs, Jack P. "The division of labor: conceptualization and related measures", *Social Forces,* 53 (3), 1975, 468-476.

Haddock, David D. "Irrelevant externality angst", Working Paper, Northwestern University, 31 August 2004.

Head, Keith and Ries, John "Immigration and Trade Creation: Econometric Evidence from Canada", *The Canadian Journal of Economics*, 31 (1), 1998, 47-62.

Kelly, Morgan. "The dynamics of Smithian growth", *The Quaterly Journal of Economics*, 112 (3), 1997. 939-964.

Kenney, Martin and von Burg, Urs "Technology, Entrepreneurship and Path Dependence: Industrial Clustering in Silicon Valley and Route 128", *Industrial and Corporate Change*, 8 (1), 67-103.

Kibritçioglu, Aykut. "On the Smithian origins of "new" trade and growth theories". *Economics Bulletin*, Vol. 2. No. 1. 2002. 1-15.

Kilbridge, Maurice and Wester, Leon. "An economic model of the division of labor", *Management Science,* 12 (6), 1966, B255-B269.

Langlois, Richard N. " Cognitive Comparative Advantage and the Organization of Work: Lessons from Herbert Simon's Vision of the Future", Department of Economics Working Paper No. 2002 -20. Storrs, CT: University of Connecticut, October 2002. January 2005: (http:// papers.ssrn.com/sol3/papers.cfm?abstract_id=353262).

Lavezzi, Andrea Mario "Division of labor and economic growth: from Adam Smith to Paul Romer and beyond", Conference: "Old and New Growth Theories: an Assessment". Pisa, 5-7 October 2001.

Levy, David. "Testing Stigler's interpretation of 'the division of labor is limited by the extent of the market", *The Journal of Industrial Economics*, 32 (3), 1984. 377-389.

Lio, Monchy "Effects of insurance and morald hazard on the division of labor and productivity", National Sun Yat-Sen University, Taiwan.

Liu, Pak-Wai and Yang, Xiaokai. "The theory of irrelevance of the size of the firm", 1999.

Lowry, S. Todd. "Recent literature on ancient Greek economic thought". *Journal of Economic Literature*, 17 (1), 1979, 65-86.

MacDonald, James M. "Market exchange or vertical integration: an empirical analysis", *The Review of Economic and Statistics*, 67 (2), 1985.327-331.

Mahoney, Michael T. "A theory of cost and intermittent production", *The Journal of Business*, 56 (2), 1983, 139-153.

McNulty, Paul J. "Adam Smith's concept of labor", *Journal of the History of Ideas*, 34 (3), 1973, 345-366.

Meek, Ronald L. and Skinner, Andrew S. "The development of Adam Smith's ideas on the division of labour", *The Economic Journal*, 83 (332), 1973, 1094-1116.

Meyer, David R. "The division of labor and the market area of manufacturing firms", *Sociological Forum* 3 (3), 1988, 433-453.

Myint, Hla "The classical view of the economic problem", *Economica*, 13 (50), 1946, 119-130.

Newman, Peter "The erosion of Marshall's Theory of Value", *The Quaterly Journal of Economics*, 74 (4), 1960, 587-599.

Ng, Yew-Kwang. "Increasing Returns and Economic Efficiency". Paper presented at the Monash Workshop on Increasing Returns and Inframarginal Economics". Clayton, VIC, Australia: Monash University, 24-28 February 2003.

Ng, Yew Kwang and Yang, Xiaokai. "Effects of externality-corrective taxation on the extent of

the market and network size division of labor",

Prendergast, Renee "Marshallian external economies", *The Economic Journal*, 103 (417), 1993, 454-458.

Rauch, James and Vitor Trindade (2002) "Ethnic Chinese Networks in International Trade", *Review of Economics and Statistics* 84:1 pp. 116-130.

Robertson, Paul L. and Alston, Lee J. "Technological choice and the organization of work in capitalist firms", *The Economic History Review*. New Series, 45 (2), 1992. 330-349.

Rosen, Sherwin. "Substitution and Division of Labor". *Economica*, 45 (179), 1978, 235-250.

Rosenberg, Nathan. "Some institutional aspects of the Wealth of Nations", *The Journal of Political Economy*, 68 (6), 1960, 557-570.

Rosenberg, Nathan. "Technological change in the machine-tool industry, 1840-1910", *The Journal of Economic History*, 3 (4), 1963, 414-443.

Rosenberg, Nathan. "Technological change in the machine-tool industry, 1840-1910", *The Journal of Economic History*, 3 (4), 1963, 414-443.

Rosenberg, Nathan. "Adam Smith on the division of labour: two views or one?" *Economica*, New Series, 32 (126), 1965, 127-139.

Rosenberg, Nathan. "Another advantage of the division of labor". *The Journal of Political Economy*, 84 (4), 1976. 861-868.

Rosenberg, Nathan. "Babbage: pioneer economist",

Saxenian, AnnaLee "Comment on Kenney and von Burg, 'Technology, Entrepreneurship and Path Dependence: Industrial Clustering in Silicon Valley and Route 128'", *Industrial and Corporate Change*, 8 (1), 105-110.

Schaffer, Simon "Babbage's intelligence: calculating engines and the factory system", *Critical Inquiry*, 21 (1), 1994, 203-227.

Schiavone, Francesco and Dezi, Luca "Division of labour, entrepreneurial sources and regional development: some empirical evidences and one theoretical proposal", 2003.

Schumpeter, Joseph; Young, Allyn A.; Dewsnup, Ernest R.; Dixon, Frank Haigh; Boyton, Arthur J.; and Haney, Lewis H. "Railway Rate Making: Discussion", *The American Economic Review*, 4 (1), 81-100.

Scott, A.J. "Industrial Organization and Location: Division of Labor, the Firm, and Spatial Process". *Economic Geography*, Vol. 62, No. 3. (Jul., 1986), pp. 215-231.

Shaw, Bernard. *Pygmalion*. New York, NY: Brentano, 1916; Bartleby.com, 1999 (accessed 1 May 2005: www.bartleby.com/138/).

Shell, Karl "Toward a theory of inventive activity and capital accumulation", *The American Economic Journal*, 62-68.

Smith, Adam "An inquiry into the nature and causes of the wealth of nations", (http://www.adamsmith.org/smith/won-b1-c3.htm).

Sraffa, Piero. "The law of returns under competitive conditions". *Economic Journal*, December 1926.

Steil, Benn, Victor, David G. and Nelson, Richard R. (eds). *Technological innovation and economic performance*, Princeton, N.J.: Princeton University Press, 2002.

Stigler, George J. "The division of labor is limited by the extent of the market". *The Journal of Political Economy*, 59 (3), 1951. 185-193.

Stigler, George J. "The successes and failures of Professor Smith", *The Journal of Political Economy*, 84 (6), 1976, 1199-1213.

Stigler, George J. *Memoirs of an unregulated economist*. New York, NY: Basic Books, 1988.

Strassmann, W. Paul "Interrelated industries and the rate of technological change", *The Review of Economic Studies*, 27 (1), 1959, 16-22.

Strassmann, W. Paul "Technological Change in the machine Tool Industry, 1840-1910: Discussion", *The Journal of economic History*, 23 (4), 1963, 444-446.

Taussig, F.W. *Principles of economics*. Vol. 1. New York, N.Y.: The MacMillan Company. 1912.

Viner, Jacob, "Adam Smith and laissez-faire", *Journal of Political Economy*, 35 (2), 1927, 207

Weld, L.D.H. "Do principles of large-scale production apply to merchandizing?" *American Economic Review*, 13 (1), 1923. 185-197.

West, E.G. "Adam Smith's two views on the division of labour". *Economica*, 31 (121), 1964, 23 -32.

White, Harrison C. *Markets from networks: socioeconomic models of production*. Princeton, NJ: Princeton University Press, 2002.

Williamson, Oliver E. *The Economic Institutions of Capitalism*. New York: Free Press. 1985.

Wright, Mike and Thompson , Steve "Vertical disintegration and the life-cycle of firms and industries". *Managerial and Decision Economics*, 7, 1986. 141-144.

Yang, Xiaokai and Ng, Siang "Specialization and the division of labor: a survey", in *Increasing returns and economic analysis*, Arrow, Kenneth and Yang, Xiaokai, eds., London: Macmillan, 1998.

Yang, Xiaokai and Yeh, Yeong-nan. "Endogenous Specialization and Endogenous Principal-agent Relationship", 2000.

Yang, Xiaokai. "Incomplete Contingent Labor Contract, Asymmetric Residual Rights and Authority, and the Theory of the Firm", 2001.

Yang, Xiaokai. *Economics: New Classical versus Neoclassical Frameworks*. Malden, MA: Blackwell Publishers. 2001.

Young, Allyn A. 1913. "Pigou's Wealth and Welfare". *Quarterly Journal of Economics* 27: 672-86.

Young, Allyn A. "Increasing Returns and Economic Progress", *The Economic Journal*, volume 38 (1928), pp. 527-42.