Market Size, Division of Labor, and Firm Productivity*

Thomas Chaney†
University of Chicago and NBER

Ralph Ossa‡
University of Chicago and NBER

July 12, 2012

Abstract

We generalize Krugman's (1979) 'new trade' model by allowing for an explicit production chain in which a range of tasks is performed sequentially by a number of specialized teams. We demonstrate that an increase in market size induces a deeper division of labor among these teams which leads to an increase in firm productivity. The paper can be thought of as a formalization of Smith's (1776) famous theorem that the division of labor is limited by the extent of the market. It also sheds light on how market size differences can limit the scope for international technology transfers.

JEL classification: F10, F12, L22, L25

Keywords: Market size; Division of labor; Firm productivity; Technology transfer

---

*We are grateful to Pol Antras, Holger Breinlich, Alejandro Cunat, Elhanan Helpman, Gianmarco Ottaviano, Henry Overman, Stephen Redding, and Tony Venables. We also thank the editor, Robert W. Staiger, and two anonymous referees, for their thoughtful comments. All remaining errors are ours. This work extends the second chapter of Ossa's Ph.D. dissertation originally entitled "Trade Liberalization, Outsourcing, and Firm Productivity".

†University of Chicago, Department of Economics, 1126 East 59th Street, Chicago, IL 60637; thomas.chaney@gmail.com

‡Corresponding author: University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637; ralph.ossa@chicagobooth.edu

Introduction

In this paper, we develop a simple general equilibrium model in which an increase in market size leads to an increase in the division of labor which brings about an increase in firm productivity. In particular, we generalize Krugman's (1979) seminal 'new trade' model by opening the black box of the production function and allowing for an explicit production chain in which a range of tasks is performed sequentially by a number of specialized production teams. An increase in market size induces a deeper division of labor among these teams which leads to an increase in firm productivity. Underlying this is a trade-off between the fixed costs associated with establishing a team and the marginal costs associated with the degree of specialization of the team which firms solve differently depending on the size of the market.

At the broadest level, the paper can be thought of as a formalization of Smith's (1776) famous theorem that the division of labor is limited by the extent of the market in an environment in which the division of labor takes the same form as in his pin factory.[1] By embedding the pin factory into a framework of monopolistic competition, it overcomes the dilemma emphasized by Stigler (1951: 185) that "either the division of labor is limited by the extent of the market, and characteristically, industries are monopolized; or industries are characteristically competitive, and the theorem is false or of little significance". An increase in market size leads to both a deeper division of labor within firms as well as the entry of new firms.

While our theory is not explicit about the nature of the increase in market size, the usual interpretation of the Krugman (1979) model suggests trade liberalization as a natural example. Recently, many empirical studies have focused on the productivity effects of trade liberalization (e.g. Pavcnik 2002; Trefler 2004). Their results suggest that there are important trade-induced improvements in industry productivity either

[1] Recall that in Smith's (1776: 7) pin factory "one man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head (...)".

through gains in average firm productivity ('firm productivity effect') or through the reallocation of market share from less to more productive firms ('reallocation effect'). While our theory cannot speak to the reallocation effect, it can be thought of as a micro-foundation of the firm productivity effect.[2]

As such, the paper contributes to a growing literature on the sources of the firm productivity effect. Previous work has mainly emphasized fixed costs (e.g. Krugman 1979), learning by exporting (e.g. Clerides, Lach, and Tybout 1998), competition-induced innovation (e.g. Aghion et al. 2005), or a horizontal focusing on core competencies by multi-product firms (e.g. Eckel and Neary, 2010; Bernard, Redding, and Schott, 2011). Only McLaren (2000) also studies the productivity gains of a trade-induced vertical restructuring of production. Both the source of the productivity gains as well as the link between trade liberalization and the vertical restructuring of production are very different in his model, however.

An additional implication of our model is that seemingly superior technologies developed in larger markets, characterized by lower fixed costs of establishing teams and a finer division of labor across teams, may not be appropriate for smaller markets. Firms in developing countries may therefore not have an incentive to adopt technologies from developed countries even if they are freely available to them. This observation offers a novel explanation for the localized character of technology which is usually rationalized by arguing that important components of technology are tacit in nature (e.g. Keller, 2004: 753). It essentially elaborates on the remark of Stigler (1951: 193) that American production methods will often be too specialized to be an appropriate model for industrialization in developing countries.

The remainder of the paper is organized as follows: we lay out the basic model, solve for the optimal organization of production, characterize the general equilibrium, analyze the effects of an increase in market size, consider the scope for international

[2]Well-known formal treatments of the reallocation effect include Melitz (2003) and Bernard et al (2003).

technology transfers, and offer some concluding remarks.

1 Basic setup

There are L consumers who are endowed with one unit of labor each. They have access to n final goods over which they have 'love of variety'-preferences

$$U = \sum_{i=1}^{n} u(x_i) \tag{1}$$

where $u(x_i)$ is the utility derived from consuming x units of final good i which is continuous and differentiable and satisfies $u'(x_i) > 0$ and $u''(x_i) < 0$. Consumers maximize this utility subject to their budget constraints $1 = \sum_{i=1}^{n} p_i x_i$, where p_i is the price paid for good i and the wage rate is normalized to 1.

As can be seen from the first order conditions of the consumers' maximization problems, the resulting demands have elasticity $\varepsilon(x_i) = -\frac{u'(x_i)}{x_i u''(x_i)}$. Following Krugman (1979), we assume that $\varepsilon'(x_i) < 0$ which is equivalent to assuming that the demand curves are less convex than in the constant elasticity case (linear demand curves would be an example). This assumption ensures that an increase in market size leads to an increase in firm output which is necessary for market size to affect the division of labor within firms. We also assume that $\varepsilon(0) > 1 + \frac{1}{\gamma}$ and that there exists an $\bar{x} > 0$ such that $\varepsilon(\bar{x}) = 1 + \frac{1}{\gamma}$, where γ is a cost parameter to be defined below.[3] These parameter restrictions guarantee the existence and uniqueness of a monopolistically competitive equilibrium.

The production of each final good requires the sequential performance of a number of tasks. Early tasks are concerned with obtaining raw materials which are then refined

[3] A polynomial of degree higher than 2 for the function $u(x)$ would satisfy this condition, as would any sum of more than one power function of x. For instance, the quadratic function $u(x) = ax - x^2/2$ with $x \in [0, a/2]$ yields a linear demand system and the following simple expression for the demand elasticity, $\varepsilon(x) = a/x - 1$, which satisfies $\varepsilon'(x) < 0$, $\varepsilon(0) > 1 + 1/\gamma$, and $\bar{x} = \frac{a}{2+1/\gamma}$ such that $\varepsilon(\bar{x}) = 1 + \frac{1}{\gamma}$.

successively in later production stages. The set of these tasks is represented by a segment of length normalized to 2 which we call the production chain. To produce the final good, all tasks $\omega \in [0,2]$ have to be performed sequentially. If only tasks $\omega \in [0,\omega_1]$, $0 < \omega_1 < 2$, are performed, a preliminary good ω_1 is obtained. This preliminary good ω_1 can then be transformed into a more downstream preliminary good ω_2, $0 < \omega_1 < \omega_2 < 2$, by performing the additional tasks $\omega \in [\omega_1,\omega_2]$ and so on. One unit of each task is required to produce one unit of the final good. Similarly, one unit of the relevant subset of tasks is required to produce one unit of a preliminary good.[4]

All production tasks associated with a given final good are performed by production teams within a single firm. Before being able to perform any tasks, a team needs to acquire a core competency $c \in [0,2]$ in the production chain which requires f units of labor. To perform one unit of each task in the range $[\omega_1,\omega_2]$, the team then further needs

$$l\left(\omega_1,\omega_2\right) = \frac{1}{2}\int_{\omega_1}^{\omega_2} |c-\omega|^{\gamma} d\omega \tag{2}$$

units of labor where $\gamma > 0$ so that it gets worse at performing a given task the further away that task is from its core competency. Teams are symmetric in the sense that the parameters f and γ are the same across teams. The firm can choose how many teams are established, which core competencies they acquire, and which production tasks they perform.

2 Optimal organization of production

Equation (2) implies that the cost of producing one unit of output is minimized if each task is performed by only one team, the teams' core competencies are uniformly distributed along the production chain, and each team performs a symmetric range of

[4] A similar representation of the production process has been used by Dixit and Grossman (1982).

tasks around its core competency. The minimum total cost of producing y units of output conditional on a given number of teams t can therefore be written as

$$TC = t\left(f + y\int_0^{\frac{1}{t}} \omega^{\gamma} d\omega\right) \tag{3}$$

since each team performs $\frac{2}{t}$ tasks of which half are to the right and half are to the left of its core competency.

The optimal number of teams solves a trade-off between fixed and marginal costs. This trade-off can be seen most clearly by rewriting equation (3) as $TC = tf + \frac{yt^{-\gamma}}{\gamma+1}$. On the one hand, more teams imply higher fixed costs since more core competencies need to be acquired. On the other hand, more teams imply lower marginal costs since each team performs a narrower range of tasks around its core competency. Minimizing this expression with respect to t yields

$$t = \left(\frac{\gamma}{\gamma+1}\frac{y}{f}\right)^{\frac{1}{\gamma+1}} \tag{4}$$

Hence, the optimal number of teams is increasing in output. Intuitively, higher output makes marginal costs more important relative to fixed costs so that it is optimal to set up a larger number of more highly specialized teams. Notice that the range of tasks performed by each team is inversely proportional to the number of teams since the production chain is of a given length and production tasks are equally divided among teams.

As is easy to verify, equations (3) and (4) imply that the average cost is given by

$$AC = \left(\frac{\gamma+1}{\gamma}\frac{f}{y}\right)^{\frac{\gamma}{\gamma+1}} \tag{5}$$

Notice that the average cost is decreasing in output so that the production technology exhibits increasing returns to scale. Underlying this are two distinct effects which can be seen most clearly by expressing the average cost as $AC = \frac{tf}{y} + t\int_0^{\frac{1}{t}} \omega^{\gamma} d\omega$ using equation (3). First, the average cost falls because the fixed costs get spread over more units of output. Second, the average cost falls because the number of teams is increased to rebalance fixed and marginal costs. Only the former effect is present in Krugman (1979).[5] The second effect magnifies the first effect since the number of teams is chosen to minimize costs.

While the details of equations (3) - (5) clearly depend on functional form assumptions, they capture what seems to be a general point: if production tasks are divided among specialized teams who need to incur a fixed cost to acquire a core competency and get worse at performing a task the further away it is from their core competency, the optimal number of teams is increasing in output since the increase in output makes marginal costs more important relative to fixed costs. We therefore state this result as proposition 1:

Proposition 1 *The optimal number of teams is increasing in firm output.*

Proof. Follows immediately from equation (4). ■

3 General equilibrium

Firms interact in a monopolistically competitive fashion in the sense that they maximize profits taking the marginal utility of income as given and enter until all profits are driven down to zero. Free entry implies that prices are equal to average costs and profit maximization implies that firms charge a proportional mark-up $\mu(x) = \frac{\varepsilon(x)}{\varepsilon(x)-1}$ over marginal costs. The equilibrium is characterized by the following two conditions,

$$p = \left(\frac{\gamma+1}{\gamma}\frac{f}{y}\right)^{\frac{\gamma}{\gamma+1}} \tag{6}$$

[5] Indeed, the model would reduce to Krugman (1979) if the number of teams were not a choice variable.

$$p = \frac{\mu(y/L)}{\gamma+1}\left(\frac{\gamma+1}{\gamma}\frac{f}{y}\right)^{\frac{\gamma}{\gamma+1}} \tag{7}$$

The first condition follows immediately from equation (5). The second condition combines the goods market clearing condition $y = Lx$ with the fact that marginal costs are given by $MC = \frac{1}{\gamma+1}\left(\frac{\gamma+1}{\gamma}\frac{f}{y}\right)^{\frac{\gamma}{\gamma+1}}$ which is obtained by straightforward manipulation of equation (5).[6] The relationships (6) and (7) are two equations in the two unknowns p and y which we refer to as the FE (free entry) curve and the PM (profit maximization) curve in the following. The FE curve is downward-sloping. The PM curve is made up of two multiplicative terms. The first term, $\frac{\mu(y/L)}{\gamma+1}$, is smaller than 1 at zero because $\varepsilon(0) > 1 + \frac{1}{\gamma}$, increasing in y because $\varepsilon'(x) < 0$, and crosses 1 at some finite $\bar{x}$ because $\varepsilon(\bar{x}) = 1 + \frac{1}{\gamma}$. The second term is the same term as the FE curve. Therefore, the PM curve intersects the FE curve only once from below, as illustrated in Figure 1, so that p and y are uniquely pinned down.[7] Given y, the equilibrium t can then be determined from equation (4).

Notice that equations (6) and (7) imply that mark-ups are constant in equilibrium even though preferences are not of the constant elasticity form. This is due to the fact that changes in the optimal division of labor ensure that marginal costs and average costs fall proportionately in firm output so that mark-ups have to be constant for zero profits to prevail. This exact proportionality of marginal costs and average costs depends on special functional form assumptions and should not be taken literally. However, it may prove useful as a modeling tool in other applications which seek to allow for a relatively general demand system without losing the tractability of constant elasticity preferences.

[6] All subscripts have been dropped to reflect the symmetry of the equilibrium.
[7] Notice that the PM curve does not have to be downward sloping.

4 Increase in market size

We first analyze an increase in market size which is captured by an increase in the number of consumers L. As can be seen from equations (6) and (7), an increase in market size leaves the FE curve unchanged but leads to a downward shift of the PM curve. Intuitively, an increase in the number of consumers implies that each consumer gets less of a given quantity of output which increases demand elasticities and reduces mark-ups, other things equal. As illustrated in Figure 2, this downward shift of the PM curve implies that firms charge less and produce more which is associated with an increase in the number of teams per firm as indicated by equation (4). Intuitively, the larger market allows firms to sell more which makes them establish a larger number of more highly specialized production teams.

Since average costs are simply the inverse of firm output per worker, the fall in average costs associated with the increase in firm output also represents an increase in firm productivity. Recall that average costs fall because the fixed costs get spread over more units of output and the number of teams is increased to rebalance fixed and marginal costs. Hence, while the model continues to feature the original Krugman (1979) firm productivity effect, it also features a new firm productivity effect which operates through an increase in the vertical division of labor. The latter effect magnifies the former effect since the degree of the vertical division of labor is chosen optimally by firms.

Hence, an increase in market size indeed leads to an increase in the division of labor which is associated with an increase in firm productivity. It must be emphasized, however, that this result depends on the fact that firm output is increasing in market size which, in turn, depends on the assumption that the demand curves are less convex than in the constant elasticity case. If utility was instead of the constant elasticity form as in Krugman (1980), the number of firms would simply increase proportionately with market size so that individual firm output would remain unchanged. With this caveat in mind, we state this result as proposition 2:

Proposition 2 *An increase in market size leads to an increase in the division of labor which is associated with an increase in firm productivity.*

Proof. Follows immediately from Figure 2 and equation (4). ■

5 Technology transfer

We now consider the scope for technology transfers from a Northern country to a Southern country, where firms in the Northern country operate seemingly superior technologies. To analyze technology transfers, we must first define what a technology is. In the context of this model, a technology has two key components. The first corresponds to the efficiency with which a firm is able to train specialized production teams around a core competency and is captured by the fixed cost f.[8] The second corresponds to the degree of division of labor within a firm and is captured by the number of specialized production teams t.

So far, we have treated only the fixed cost f as a parameter and allowed firms to optimally choose their organization t. This endogenous choice was meant to capture a long-run adjustment during which incumbents either reorganize or lose out to better organized entrants. It is plausible, however, that both the fixed cost f and the organization t have to be jointly transferred in the case of international technology transfers. This is because the organization of production solves a complex logistical problem so that a Southern firm is unlikely to be able to rearrange the production chain of a Northern firm to appropriately reflect local constraints.

Given the premise that technology transfers entail both f and t, it is now easy to see that a Southern firm might be unwilling to adopt a seemingly superior Northern

[8] Of course, the efficiency with which a firm is able to train specialized production teams around a core competency also depends on the parameter γ. However, our point can be made most clearly with reference to the parameter f so that we focus on it in the following.

technology. In particular, suppose that the North has a larger market ($L_N > L_S$) and lower fixed costs ($f_N < f_S$) so that the division of labor is finer there ($t_N > t_S$). If a Southern firm then adopts the Northern technology, its profitability increases on the one hand due to the lower fixed cost but decreases on the other hand as a result of the suboptimally large number of teams. The former effect dominates the latter one only if the market size differences underlying the differences in the optimal organization of production are sufficiently small.

As a result, international technology transfers may not occur even though Northern firms are unambiguously more productive and make their technology freely available to Southern firms. We believe that this offers a novel perspective on the notion of appropriate technology (e.g. Basu and Weil, 1998) and a novel explanation for the localized character of technology which is usually rationalized by arguing that important components of technology are tacit in nature (e.g. Keller, 2004: 753). We state our reasoning more rigorously as proposition 3:

Proposition 3 *For any difference in market size between a larger North and a smaller South ($L_N > L_S$), there exists $\Delta > 0$ such that no Southern firm would want to adopt a Northern technology characterized by the Northern vertical division of labor t_N unless it offers them a reduction in their fixed cost at least as large as Δ.*

Proof. Proposition 2 and equation (4) imply that $t_N > t_S$. From equation (3) and the optimal choice of t given f, it follows directly that a Southern firm's maximum profits, π_S, are decreasing in f and decreasing with departures away from the optimal t_S: π_S satisfies $\pi_S(f_S, t_S) > \pi_S(f_S, t_N)$, $\frac{\partial \pi_S}{\partial f} < 0$, $\frac{\partial \pi_S}{\partial t}(f_S, t_S) = 0$, and $\frac{\partial^2 \pi_S}{\partial t^2} < 0$. Therefore there exists a $\Delta > 0$ such that $\pi_S(f_S, t_S) = \pi_S(f_S - \Delta, t_N)$. For any reduction in the fixed cost smaller than Δ, i.e. $f_S - \Delta < f_N < f_S$, we have $\pi_S(f_S, t_S) > \pi_S(f_N, t_N)$, and no Southern firm would adopt the Northern technology (f_N, t_N) despite its strictly lower fixed cost. ■

Conclusion

"As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or, in other words, by the extent of the market." In this paper, we have demonstrated that this famous theorem of Smith (1776: 16) can be rationalized by embedding a production chain of the sort found in his pin factory into Krugman's (1979) seminal 'new trade' environment. In a nutshell, we first established that the division of labor is limited by the extent of firm output and then demonstrated that firm output is increasing in the extent of the market. We also showed that in such an environment, seemingly superior technologies developed in large markets may not be appropriate for smaller markets thus limiting the scope for international technology transfers.

References

[1] Aghion, P., Burgess, R., Redding, S., Zilibotti, F, 2005. Entry liberalization and inequality in industrial performance. Journal of the European Economic Association 3, 291-302.

[2] Basu, S., and D. Weil. 1998. Appropriate technology and growth. Quarterly Journal of Economics 113(4): 1025-1054.

[3] Bernard, A., Eaton, J., Jensen, J., Kortum, S., 2003. Plants and productivity in international trade. American Economic Review 93, 1268-1290.

[4] Bernard, A., Redding, S., Schott, P., 2011. Multi-product firms and trade liberalization. Quarterly Journal of Economics, forthcoming.

[5] Clerides, S. K., Lach, S., Tybout, J.R., 1998. Is learning by exporting important? Micro-dynamic evidence from Columbia, Mexico, and Morocco. Quarterly Journal of Economics 113, 903-947.

[6] Eckel, K., Neary, J.P., 2010. Multi-product firms and flexible manufacturing in the global economy. Review of Economic Studies 77, 188-217.

[7] Dixit, A., Grossman, G. M., 1983. Trade and protection with multi-stage production. Review of Economic Studies 49, 583-594.

[8] Keller, W., 2004. International Technology Diffusion. Journal of Economic Literature 42: 752-782.

[9] Krugman, P., 1979. Increasing returns, monopolistic competition, and international trade. Journal of International Economics 9, 469-479.

[10] Krugman, P. 1980. Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review* 70(5): 950-959.

[11] Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695-1725.

[12] McLaren, J., 2000. Globalization and vertical structure. American Economic Review 90, 1239-1254.

[13] Pavcnik, N., 2002. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. Review of Economic Studies 69, 245-276.

[14] Smith, A., 1776. *An inquiry into the nature and causes of the wealth of nations.* Volume 1. London, UK: Strahan and Cadell.

[15] Stigler, George J., 1951. The Division of Labor is Limited by the Extent of the Market. Journal of Political Economy 59, 185-193.

[16] Trefler, D., 2004. The long and short of the Canada-U.S. free trade agreement. American Economic Review 94, 870-895.

[17] Tybout, J., 2003. Plant- and firm-level evidence on the "new" trade theories. In: Kwan Choi, E., Harrigan, J. (Eds.), Handbook of International Trade, Basil-Blackwell, Oxford.

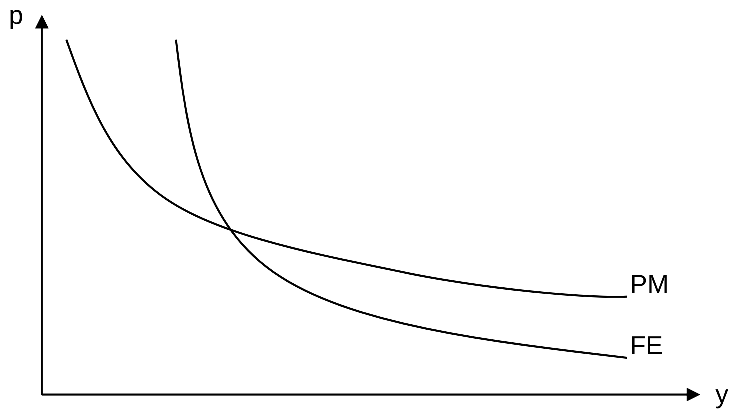

Figure 1: General equilibrium

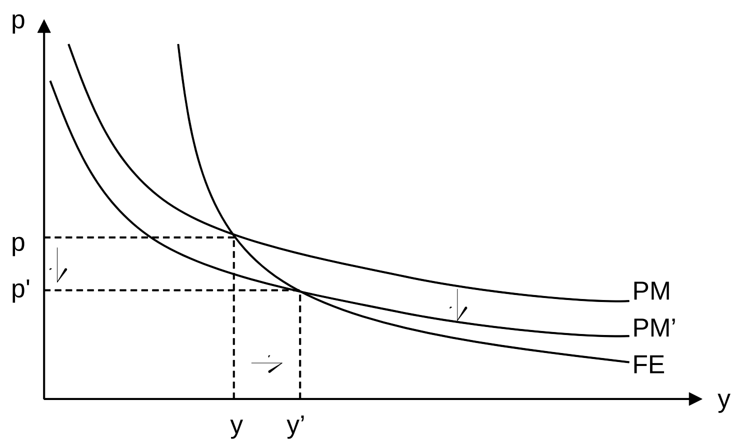

Figure 2: Increase in market size