MONOPOLISTIC COMPETITION
IN TRADE THEORY

ELHANAN HELPMAN
SPECIAL PAPERS IN INTERNATIONAL ECONOMICS

SPECIAL PAPERS IN INTERNATIONAL ECONOMICS are published by the International Finance Section of the Department of Economics of Princeton University. While the Section sponsors the Special Papers, the authors are free to develop their topics as they wish. The Section welcomes the submission of manuscripts for publication in this and its other series. See the Notice to Contributors at the back of this publication.

The author of this Special Paper, Elhanan Helpman, is the Archie Sherman Professor of International Economic Relations at Tel Aviv University. His publications include A Theory of International Trade under Uncertainty (with Assaf Razin), Market Structure and Foreign Trade, and Trade Policy and Market Structure (with Paul R. Krugman).

This paper was prepared while the author was a Fellow at the Institute for Advanced Studies, The Hebrew University of Jerusalem, and was presented as the Frank Graham Memorial Lecture at Princeton University on April 24, 1989.

PETER B. KENEN, Director
International Finance Section
MONOPOLISTIC COMPETITION
IN TRADE THEORY

ELHANAN HELPMAN
CONTENTS

1 INTRODUCTION 1

2 INTRA-INDUSTRY TRADE 3

3 APPLICATIONS 7
   Trade Volume 7
   Share of Intra-Industry Trade 8
   North-South Trade 9
   Factor Movements 10

4 TARIFFS 13
   Terms of Trade 13
   Production Efficiency 14
   Variety Choice 16

5 MULTINATIONAL CORPORATIONS 20

6 TRADE DYNAMICS 23

7 LONG-RUN GROWTH 28
   Endogenous Growth 28
   The North-South Product Cycle 30
   Policy 33

8 CONCLUDING COMMENTS 36

REFERENCES 37
**LIST OF FIGURES**

1. Tariffs and Consumer Price Indexes 18
2. Consumer Price Indexes and Variety 19
3. Multinational Corporations 21
4. Trade Dynamics 25
5. North-South Product Cycle 32
1 INTRODUCTION

In 1979, Kelvin Lancaster and Paul Krugman published independent formalizations of an idea that had been around for many years, namely, that the manufacture of differentiated products with brand-specific economies of scale leads to intra-industry trade (two-way trade in similar, although not necessarily identical, products). They made their point with simplified one-sector models in which all trade is of the intra-industry type. Though they used different approaches to the specification of preferences and other details, the same central message emerged from their writings: the time was ripe for an incorporation of important sectors of the industrial world into the formal theory of international trade (see also Balassa, 1967; Grubel and Lloyd, 1975; and Norman, 1976).

In the ensuing years, this building block was effectively used to reformulate trade theory. It also opened the door to a broader treatment of non-competitive market structures. Thus, for example, Frank Graham's famous argument for tariff protection and his debate with Frank Knight (see Knight, 1924, 1925, and Graham, 1925) were examined with modern tools. It was shown that Graham was right (see Ethier, 1982b): a country that produces import-competing goods with increasing returns to scale may lose from trade, and a tariff may help in these circumstances.

Ten years after the turning point is a suitable time to take stock of these developments. The entire literature on noncompetitive trade theory is too vast to be reviewed in a single paper, so I confine the discussion to monopolistic competition in differentiated products. Restrictive as this choice may seem, it has much to offer. Not only was this line of research central to the development of the new theory of international trade during the 1980s, but it also has become central to the recent rethinking of macroeconomics in general and economic growth in particular. And, most recently, it has become a cornerstone in the treatment of dynamic trade issues.

I discuss substantive issues in two parts. The first part begins with a brief review of developments in the early 1980s, emphasizing fundamentals (Chapter 2). Then I show how the basic framework was applied to various problems, such as the explanation of the volume of trade and the share of intra-industry trade, the effect of resource expansion on North-South terms.
of trade and welfare, and the effects of international factor mobility (Chapter 3). In Chapter 4 I explain commercial-policy implications in a static framework. The first part closes with a discussion of multinational corporations (Chapter 5).

The second part deals with dynamic issues. Chapter 6 begins with the description of a recent formalization of endogenous product innovation. This approach is then used to characterize the dynamic evolution of trade. In Chapter 7 the approach is used to describe endogenous product cycles, the link between endogenous long-run growth and structural features of the international economy, the effects of commercial and industrial policies on long-run growth, and the relationship between growth promotion and economic welfare.
INTRA-INDUSTRY TRADE

Lancaster (1979, Chap. 10) and Krugman (1979a) designed their work to describe intra-industry trade. They formalized an economic story that can be summed up as follows: Certain industries manufacture many varieties of the same product. Producers cater to markets in which there is a demand for a wide spectrum of brands. To penetrate the market with a new brand, the manufacturer must incur fixed costs arising from the need to develop, advertise, and market the product. Nevertheless, the existence of brand-protection rights and the economic calculus itself suggest that entrepreneurs will find it profitable to differentiate their products from those of other suppliers. Therefore, every manufacturer ends up supplying a different brand. More specifically, in an integrated world market every country specializes in a subset of the available brands.

Once this is understood, the next step is straightforward. Suppose there is a demand in every country for a wide spectrum of brands. It may arise from consumers' varied tastes for final goods or from producers' demand for differentiated intermediate inputs. Because every country specializes in a different subset of brands, it will import brands that are not produced at home, thereby bringing about intra-industry trade.

In the early 1980s, this idea was formally incorporated into multisector models by Dixit and Norman (1980, Chap. 9), Lancaster (1980), Helpman (1981), and Ethier (1982a). These extensions were important because they allowed a clear distinction between intersectoral and intra-industry trade, a distinction that did not exist in the 1979 formulations. Every contribution used a blend of Chamberlin's (1933) notion of monopolistic competition in horizontally differentiated products (the large-group case) and a formal structure of preferences that relied either on the love-of-variety approach proposed by Dixit and Stiglitz (1977) or on the ideal-variety approach proposed by Lancaster (1979). Alternative specifications of preferences did not make much difference, however, as far as trade structure was concerned. The critical element was the preferences' ability to provide brand-specific demand functions and, from them, brand-specific elasticities of demand.

In order to identify the basic elements of this approach, it is easiest to examine first a single, fully integrated world economy in which technology is the same everywhere and factor inputs move freely around the globe. Also assume that in an industry capable of manufacturing different brands all brand-specific production functions are the same. Profit-maximizing producers equate marginal revenue to marginal costs, leading to a symmetrical
equilibrium in which all brands of a given industry are supplied in the same quantity and equally priced:

\[ p_i = R_i(p,n)mc_i(w,x_i) \]  

(1)

where \( i \) is an industry index; \( p_i \) is the price of a product in industry \( i \); \( R_i(\cdot) \) is the markup of price over marginal costs, which depends on the elasticity of demand; \( mc_i(\cdot) \) is marginal costs; \( p \) is the vector of product prices; \( w \) is the vector of primary input prices (intermediate inputs are assumed away at this stage); \( x_i \) is output per brand in industry \( i \); and \( n \) is the vector whose typical element is the number of brands in industry \( i \), denoted by \( n_i \). With constant returns to scale, marginal costs are independent of output; with perfect competition, the markup function is identically equal to 1. Under these conditions, equation (1) reduces to the standard output-independent pricing condition: price equals marginal cost. With the Dixit-Stiglitz specification of preferences, the markup function is a constant larger than 1, while with Lancaster’s specification, it depends on prices and the number of brands.¹

Following Chamberlin, assume that free entry drives profits down to zero (the large-group case). Then price equals average (unit) costs:²

\[ p_i = c_i(w,x_i) \]  

(2)

Unit costs \( (c_i) \) decline with output whenever there are increasing returns to scale.

The pricing equations (1) and (2), the former resulting from profit maximization and the latter from free entry, coincide for competitive constant-return sectors but not for sectors with increasing returns to scale that supply differentiated products. Given factor rewards and the number of brands, they determine prices and output per brand. The result is that employment of an input is the same for every brand in a given sector. Hence, if \( a_i(w,x_i) \) is the vector of employment per unit of output and \( X_i (= n_ix_i) \) is aggregate output in industry \( i \), factor-market clearing requires that

\[ V = \sum a_i(w,x_i)X_i \]  

(3)

where \( V \) is the vector of available inputs.³

The model is closed with a specification of product market-clearing conditions of the usual type (see Helpman and Krugman, 1985, Chap. 7, for an explicit statement). The point is that the entire system can be used to solve for prices, factor rewards, the number of brands in every sector, output per

¹ In fact, in the Dixit-Stiglitz specification, the markup function depends on the number of brands in the industry unless there is a continuum of brands.

² Unit costs \( c_i(w,x_i) \) are related to marginal costs by \( mc_i(w,x_i) = c_i(w,x_i) + x_ic_i(w,x_i) \).

³ The unit-output employment vector \( a_i(w,x_i) \) equals the gradient of the unit-cost function \( c_i(w,x_i) \) with respect to \( w \).
brand in every sector, and the sectoral allocation of inputs. Now one can ask
two questions: If the world is divided into countries by dividing the input
vector $V$ into country-specific inputs, (1) are there world structures for
which international trade leads to an equilibrium with the essential features
of an integrated world, and (2) what is the nature of trade in such equilibria?
These questions were addressed by the factor-proportions theory, and they
were carefully investigated for many years (see Travis, 1964, Chap. 2, and
Dixit and Norman, 1980, Chap. 4). For this reason, answering them in the
extended framework provides a natural way of discovering the value added
by the new approach.

The answer to the first question is in the affirmative. Moreover, the charac-
terization of the relevant set of world structures follows step by step the anal-
ysis of competitive constant-return economies, except for one little twist.
Recall that in the traditional framework the set of world structures that
ensures factor-price equalization is constructed by adding up all possible
cross-country distributions of the sectoral employment vectors $a_i(\cdot)X_i$. This
way, every country can produce part of the aggregate-output vector with
the same techniques of production that are employed in the integrated equi-
librium, ensuring an aggregate level of world output that equals the level of
output in the integrated equilibrium. Given identical homothetic prefer-
ences, or given that every owner of inputs is located in the same country as
his inputs, this ensures market clearing at the original commodity prices and
factor rewards. The same argument applies when some sectors produce dif-
ferentiated products with brand-specific economies of scale, except that the
distribution of those sectoral employment vectors has to be restricted to
multiples of the firm-specific employment vectors $a_i(\cdot)x_i$. This restriction is
of no consequence whenever the market provides a continuum of brands.
Otherwise, the set of structures providing factor-price equalization is much
smaller (see Helpman and Krugman, 1985, Chap. 7). A critical feature that
allows us to reproduce this result is that in the integrated equilibrium all
brands of the same good are manufactured with identical inputs per unit of
output.

This feature also implies that Vanek’s (1968) chain proposition holds. Each
country is a net exporter of the services of those inputs with which it is
relatively well endowed. Here, too, one can use the standard argument.
The factor content of a country’s net imports equals the difference between
the factor contents of consumption and production. The former equals a
share of the world’s endowment of inputs, where the share is the country’s
share in world spending. This stems from preference homotheticity. The
latter equals the country’s input vector. Hence, we obtain the well-known
relationship between the international flows of factor content and the factor
endowments, thereby answering part of the second question.

With differentiated products, however, a reproduction of the integrated
equilibrium also requires the correct number of brands of every product. Hence, it is not enough to endow each country with inputs that are multiples of brand employment vectors; it is also necessary to ensure that the inputs are used to manufacture the correct number of brands. This implies that countries have to specialize in different brands. Since all brands are demanded in each and every country, we have intra-industry trade. We measure the extent of intra-industry trade between two countries, say $k$ and $j$, in a particular product $i$ by twice the minimum of the bilateral exports of product $i$. For a differentiated product this is given by $2 \min \{s^k p_i X_i, s^j p_i X_i^j\}$, where $s^k$ and $s^j$ represent the shares of countries $k$ and $j$ in world spending. Country $k$ imports from $j$ its pro rata share of $j$'s output of each and every brand, and country $j$ imports its pro rata share of $k$'s output.
We have seen that the more general theory preserves some fundamental features of the neoclassical approach. This is quite remarkable given the introduction of economies of scale and imperfections in market structure. If all one could achieve were a reproduction of neoclassical results, however, the usefulness of these generalizations would be severely limited. Their main power comes from their ability to shed new light on old questions and to handle new problems. I will give four examples of this ability.

**Trade Volume**

The factor-proportions theory predicts larger trade volumes the larger the difference in the relative composition of factor endowments of the trading partners. This stems from the fact that trade is driven by differences in factor composition (as measured by relative ratios). In their absence, there is no trade. At the same time, this theory makes no prediction concerning the role of differences in country size in determining the volume of trade. In practice, however, there are large trade volumes between countries with similar factor proportions, and relative country size seems to play an important role in explaining them (see Linnemann, 1966).

The more general approach predicts a link between the volume of trade and differences in factor proportions when some sectors supply homogeneous products, precisely because such a link exists in the factor-proportions theory (see Helpman, 1981). In the more general approach, however, intra-industry specialization also drives trade, so that it can explain trade flows between countries with similar factor proportions. In addition, it assigns a natural role to relative country size.

In order to see the latter point as sharply as possible, consider a world in which all sectors manufacture differentiated products and sectoral preferences are homothetic. In this world, the share \( s^k \) of country \( k \) in world spending likewise defines the share of that country's imports from \( j \) of every brand manufactured in \( j \) and vice versa. Hence, \( k \)'s imports from \( j \) equal a proportion \( s^j \) of \( j \)'s GDP, denoted by \( G^j \). Assuming that expenditure is proportional to GDP, the bilateral volume of trade is given by

\[
T_{ij} = s^i G^j + s^j G^i = 2G^j G^i / G ,
\]  

where \( G \) is world GDP. Hence, in a cross-country comparison we should observe trade between countries with similar factor proportions, and bilat-
eral volumes of trade should be positively related to income levels. Both predictions conform to the evidence. Moreover, equation (4) yields the following formula for the world’s volume of trade:

\[ T = [1 - \Sigma_k (s^k)^2] G , \]

where the term in the square brackets measures the degree of dispersion in relative country size. It therefore states that trade as a proportion of GDP is larger the more similar countries are in size. In the post-war period, differences in relative country size have declined while trade has grown faster than GDP, as this formula would indeed predict (see Helpman, 1987).

The critical element in this analysis is the degree of specialization. The analysis shows that high degrees of specialization assign an important role to relative country size in the determination of the volume of trade, while monopolistic competition in differentiated products leads naturally to high degrees of specialization of the intra-industry type. It is not intersectoral specialization, as in Ricardian models, but specialization nevertheless. This tendency toward intra-industry specialization has additional implications, as I show next.

**Share of Intra-Industry Trade**

To a large extent, the impetus for the new line of research came from a desire to explain intra-industry trade, and the extended model can indeed be used to decompose the total volume of trade into intra-industry and intersectoral. It can therefore be used to investigate the determinants of the share of intra-industry trade. For this purpose consider a simple world with two inputs, two sectors, and two countries. Also assume that factor-price equalization obtains. Then condition (3) implies that every country produces relatively more of the good that is intensive in the input with which it is relatively well endowed (which is the Rybczynski effect). Let both sectors manufacture differentiated products. In this case, two-way trade prevails in both sectors, but every country is a net exporter of products that are relatively intensive in the input with which it is relatively well endowed. This, of course, is the Heckscher-Ohlin intersectoral pattern of trade.

Using the formula for the measurement of intra-industry trade that was derived in Chapter 2, the volume of intra-industry trade in this simplified world can be represented by

\[ T_I = 2\min\{s^1p_1X_1^1, s^2p_1X_1^2\} + 2\min\{s^1p_2X_2^1, s^2p_2X_2^2\} . \]

Now denote by \( \theta_m^k, k = 1,2 \), the GDP share of country \( k \)'s import-competitive sector (hence, \( \theta_m^k < 1 - \theta_m^k \)). Then, using equation (4), this equation can be rewritten as

\[ T_I = (\theta_m^1 + \theta_m^2)T . \]
That is, the share of intra-industry trade equals the GDP share of country 1’s import-competing sector plus the GDP share of country 2’s import-competing sector. In the limiting case in which both countries have the same composition of relative inputs, the share of intra-industry trade equals 1, that is, there is no intersectoral trade. The larger the difference in factor proportions, the smaller are the shares of the import-competing sectors in GDP and the share of intra-industry trade. When both countries specialize in the exporting sector, the share of intra-industry trade equals 0, that is, all trade is intersectoral. Hence, this model predicts smaller shares of intra-industry trade for countries with larger differences in factor proportions (see Helpman, 1981). Numerous empirical studies support this prediction (e.g. Balassa, 1986, and Helpman, 1987).

North-South Trade

There are many facets to the argument that, absent explicit policies, the secular worsening of the South’s terms of trade is inevitable and that, as a consequence, not only must its relative position decline but the standard of living of its residents also must decline. Two elements seem to play an important role in this line of reasoning: (1) the North exports manufactures while the South exports raw materials, and (2) the North exercises monopoly power. Dixit (1984) examined this issue in a framework that contains both elements by postulating that the North produces differentiated products from inputs that are imported from the South.

In order to see the importance of product differentiation in this argument, consider a stripped-down version of his model. Every country consumes only differentiated products that are produced in the North. Preferences are of the symmetric CES type, with the elasticity of substitution given by \( \sigma = \frac{1}{(1 - \alpha)} > 1 \). The South, which is competitive, produces only one good, an input that is required in the production of Northern manufacturers. One unit of Southern labor produces a unit of this input, and one unit of this input is needed to produce a unit of any variety of the differentiated product. Hence, the price of the input \( p_e \) equals the South’s wage rate \( w_s \). Given the market power of Northern manufacturers, however, they mark up price above marginal costs. Their marginal costs equal the price of intermediates \( p_e \), while their elasticity of demand equals \( \sigma \). Therefore \( \alpha p = p_e = w_s \), where \( p \) is the price of a variety of the differentiated product. The South’s terms of trade are thus fixed at \( \frac{p_e}{p} = \alpha \) and aggregate output of intermediate goods equals the South’s labor force \( L_s \).

Northern producers need to hire \( f \) units of Northern labor in order to produce a brand. This input requirement generates fixed costs. There is free entry into the industry. Therefore, the number of products \( (n) \) is \( \frac{L_N}{f} \), where \( L_N \) is the North’s labor force, and price equals unit costs. This condition, together with the previous pricing equations, implies
\[ \frac{w_N}{w_S} = (1 - \alpha)\frac{L_S}{L_N}. \]

Hence, relative wages are inversely related to the relative size of the labor force.

From the CES utility function, we find that the welfare level of a typical worker in country \( i \) is proportional to \( (w_i/p)^{\eta (1 - \alpha)/\alpha} \). Therefore, given that \( \alpha p = w_s \), a Southern worker’s welfare depends only on the number of products. By implication, a Southern worker prefers a larger North but is indifferent to the size of the South. This feature underlines the importance of product variety. In particular, it shows that Southern workers can gain from an expansion of the North even when expansion does not affect commodity terms of trade, because they prefer more variety choice.

In a more elaborate model that allows for substitution between Northern labor and imported inputs in fixed and variable costs, labor growth in the South leads to a deterioration of its terms of trade. But it also leads to an increase in variety. The former is detrimental to a Southern worker’s welfare, while the latter is helpful. The variety effect dominates as long as the elasticity of substitution in these cost components is sufficiently high (see Dixit, 1984). This shows that variety effects can be as important as terms-of-trade effects. A similar point is made by Krugman (1981), who has shown that a factor of production that is hurt by the Stolper-Samuelson proposition, in the sense that the purchasing power of its earnings is reduced, may nevertheless gain in welfare terms if the change producing the fall in its purchasing power also expands variety choice.

**Factor Movements**

My last example concerns the role of product differentiation in the analysis of factor movements. In a competitive economy with nonincreasing returns to scale, GDP depends on commodity prices and factor endowments; that is, \( G = G(p,V) \), where \( G(\cdot) \) represents the maximum value of output that can be achieved at the price vector \( p \) with the available technology and factor endowments. An important property of such economies is that the contribution to \( G(\cdot) \) of a marginal unit of an input exactly equals its market reward. This implies that a small country facing constant commodity prices and constant rewards to internationally mobile factors of production need not adopt policies to encourage or discourage either trade or factor movements; the private calculus coincides with the social calculus. If the domestic reward to a factor of production falls short of the international reward, private incentives lead to exports of the input, which increases home GNP (in this context, GNP equals GDP plus earnings of domestic inputs abroad minus earnings of foreign inputs at home). Conversely, if the domestic reward exceeds the international reward, private incentives lead
to imports of the input, which also increases home GNP. When all these adjustments are completed, and home rewards to internationally mobile factors of production equal international rewards, it is impossible to increase GNP further, and welfare reaches the highest achievable level.

This elegant result does not apply to economies with monopolistic competition in differentiated products, for two reasons (see Helpman and Razin, 1983). First, market forces do not lead to the highest value of GNP. Second, as we saw in the previous example, welfare depends not only on real income as usually measured but also on variety choice, and market forces will not necessarily provide the most desirable variety choice.

In order to see the importance of the first point, observe that if we treat as constant the output levels per brand $x_i$, then equations (2) and (3) represent the standard competitive-pricing condition and factor-market clearing condition that ensure the highest GDP level for given values of $x_i$. In fact, this system looks very much like the production system with technological parameters $x_i$ that was discussed in Jones's (1965) classic paper, and I will come back to this analogy in a moment. Assume for simplicity that only sector 1 manufactures differentiated products, while the others produce homogeneous products with constant returns to scale. Then, given the level of $x_i$, the economy's GDP can be represented by the function $G(p,V,x_1)$; this is the highest value of output that can be attained with the available technology and factor inputs, given the prices $p$ and output per brand $x_1$ in sector 1. Moreover, if $p$ and $x_1$ are equilibrium values, then equilibrium factor rewards equal the gradient of $G(\cdot)$ with respect to $V$.

Now the analogy with technology parameters proves most useful. Owing to increasing returns to scale, an increase in output per brand reduces unit costs in sector 1. It therefore acts as technical progress on the value of output. Hence, $G(\cdot)$ increases with $x_1$. It is, in fact, easy to show that

$$\frac{\partial G(\cdot)}{\partial x_1} = p_i n_i \epsilon_1 ,$$

where $\epsilon_1$ equals minus the elasticity of the unit-cost function $c_i(\cdot)$ with respect to output. Using this result and the fact that factor rewards equal the gradient of the GDP function with respect to inputs, we obtain a formula for the marginal contribution of input $h$ to GDP:

$$w_{h} = w_h + p_i n_i \epsilon_1 d x_1 / dV_h ,$$

where $w_{h} = dG(\cdot)/dV_h$, and $w_h [ = \partial G(\cdot)/\partial V_h]$ represents the domestic factor reward. It is clear from this formula that the market reward to factor $h$ underestimates its marginal contribution to the value of output if a larger supply of that factor raises output per brand in the industry producing differentiated products, and it overestimates the contribution to GDP if a larger supply of that factor reduces output per firm in that industry.
Helpman and Razin (1983) showed that in some circumstances an addition to the supply of an input increases output per brand and in other circumstances it reduces output per brand. Hence, domestic factor rewards may underestimate or overestimate the marginal contribution of an input to GDP. This implies that, even in a smaller country facing given world prices and rewards to internationally mobile factors of production, private decisions about factor movements—those based on a comparison of domestic and international factor rewards—do not lead to the highest GNP level. For example, domestic owners of capital may choose to invest abroad because the international rental exceeds the domestic rental, while at the same time the marginal contribution of capital to domestic GDP exceeds the international rental. This demonstrates that the national cost-benefit analysis of factor movements can be complex even if the problem of variety choice is totally disregarded.

In a proper cost-benefit analysis, one cannot disregard the effects of factor movements on variety choice. In order to see what may be involved, consider a case in which the differentiated products are not traded internationally (think about services such as restaurant meals or theater shows in Hebrew). Then the contribution of an imported input to home welfare is measured not only by its contribution to GNP but also by the social value of the change in variety choice that it brings about. If, for example, the international reward to \( h \) falls short of the domestic reward and an additional unit of \( h \) increases the variety choice, then private incentives to import the input also raise national welfare. In fact, private agents will not import enough of it. However, not all external effects will work in the same direction. An additional unit of \( h \) may, for instance, increase output per brand but reduce the variety choice. If the latter effect is strong enough, the private incentive to import the input will work against social welfare. This demonstrates how much sophistication is required for a precise analysis of the desirability of factor movements, in sharp contrast to the clean competitive case.
If my presentation of the North-South trade and factor-mobility applications has been clear enough, the reader should be convinced by now that every policy analysis for economies with monopolistic competition in differentiated products has to deal with the policy's effect on output per brand and the number of brands produced (variety choice). These are indeed two novel features with which an evaluation of tariffs has to cope. But an additional aspect of commercial policy deserves our attention. The traditional argument for tariffs that improve the terms of trade can be more forceful in the presence of monopolistic competition in differentiated products, because every brand faces a downward-sloping demand curve. The most remarkable implication of this fact is that even small countries can improve their terms of trade by means of a tariff. After discussing this point, I turn to the role of output per brand and variety choice.

Terms of Trade

In order to concentrate on the terms of trade, it is best to consider economies in which tariffs do not affect variety and output per brand. Suppose that in a two-country world each country uses only labor to manufacture a single differentiated product. The labor requirement per brand (the inverse of the production function) is \( g(x) \), where \( x \) represents the brand's output. Labor input per unit of output declines with the output level. In addition, let preferences be of the CES type with an elasticity of substitution \( \alpha = 1/(1 - \alpha) > 1 \). Then, as in the North-South example, the markup factor in equation (1) is constant: \( R = 1/\alpha \), and the pricing equations (1) and (2) for country \( k \), \( k = 1,2 \), imply \( \alpha p^k = w^k g'(x^k) \) and \( p^k = w^k g(x^k)/x^k \). These pricing equations uniquely determine the price/wage ratio \( p^k/w^k \) and output per brand \( x \). They are the same for both countries. It is quite clear from this exposition that output per brand does not depend on the tariff rate. In addition, given the output level of a typical brand \( x \), full employment of labor requires \( n^k g(x) = L^k \). Therefore, the number of brands produced in country \( k \), which is proportional to the size of its labor force, is also independent of the tariff rate. Consequently, the tariff can achieve only one thing: it can change relative wages and the terms of trade.

Country 1 can use an import tariff to improve its terms of trade. The optimal tariff, shown by Gros (1987), can be derived by a method borrowed from Helpman and Krugman (1989, Chap. 7). First observe that an \textit{ad valorem} tariff on all imports does not affect the relative prices of brands that
are manufactured in country 1. Therefore, we can form a Hicks aggregate of country 1's output. In the same way, we can form a Hicks aggregate of country 2's output. Having done this, we can follow the standard analysis of an optimal tariff, because we are now dealing with a world in which every country specializes in the production of a different good (a different Hicks aggregate). The optimal tariff rate equals $1/(\ell^2 - 1)$, where $\ell^2$ represents country 2's import elasticity of demand (defined to be positive). For the symmetric CES demand structure, $\ell^2 = (1 - \gamma^2) + \gamma^2\sigma$, where $\gamma^2$ is country 2's expenditure share on its own products. Therefore, we obtain the optimal tariff formula for country 1: $\tau^1 = 1/\gamma^2(\sigma - 1)$. As expected, the larger country 2 is relative to 1, the larger its expenditure share on its own products and the smaller country 1's optimal tariff rate. The striking result, however, is that even when country 2 grows infinitely larger than 1, the optimal tariff does not converge to 0 but rather to $1/(\sigma - 1)$. Hence, even a small country's optimal tariff rate is strictly positive.

There are two ways to explain this result. First, note that with CES preferences, relative country size does not affect the elasticity of demand for a single brand and every country specializes in a different range of products. Therefore, even when a country's relative size is negligible, it nevertheless maintains monopoly power in the range of products in which it specializes (to be precise, one needs to assume that there is a continuum of brands). Second, in a relatively small country that treats import prices as given, marginal import costs equal import prices. Therefore, imports are priced according to marginal cost, while domestic products are priced above marginal cost. This relative-price distortion can be corrected with a tariff. Indeed, the optimal tariff rate $1/(\sigma - 1)$ brings about equality of relative prices and relative marginal costs.

Production Efficiency

Besides affecting the terms of trade, tariffs can affect the degree of efficiency in production and variety choice. If variety choice is constant, however, we can measure the welfare change ($dU$) that results from tariffs with an equation developed by Helpman and Krugman (1989; Chap. 2):

$$dU = -m \cdot dp^* + t \cdot dm + (p - c) \cdot dX,$$

where $m$ is the vector of imports and exports (positive entries represent imports and negative entries represent exports), $p^*$ is the vector of international prices, $t$ is the tariff vector (an entry equals the tariff rate times the foreign import price), $p$ is the vector of domestic prices, $c$ is the vector of marginal production costs, and $X$ is the output vector ($X_i = n_i x_i$ with $n_i$ constant). The first term represents the usual terms-of-trade effect, and the
second term represents the usual tariff-distortion effect. The last term represents the production-efficiency effect: welfare increases if industries that price above marginal costs expand output. Output changes in competitive sectors that price according to marginal costs have no first-order effects. The last term was nil in the previous section, and the optimal tariff traded off the gain from the first term against the loss from the second.

In order to isolate the production-efficiency effect, we examine an economy in which variety choice and the terms of trade do not respond to tariffs. In such an economy, small tariffs (for which the second term is nil) affect welfare only through changes in output per brand (but see Flam and Helpman, 1987, for the role of interactions). The point I wish to make is that tariffs on competing foreign brands do not necessarily ensure longer production runs for domestic products.

As before, consider a world consisting of two economies manufacturing differentiated products that enter a CES preference function symmetrically. Now suppose, however, that countries also supply competitively a homogeneous consumption good that is produced with a unit of unskilled labor per unit of output. Take the homogeneous good to be the numeraire and consider equilibria in which both countries produce it. Then the wage rate of unskilled workers equals 1 ($w^k = 1$, $k = 1, 2$, where $L$ now stands for unskilled labor).

The production of a brand of the differentiated product requires $f$ units of skilled labor (in order, say, to develop the brand) plus $\alpha$ units of unskilled labor per unit of output (recall that the markup factor of price over marginal costs equal $1/\alpha$). Therefore, marginal costs equal $\alpha$, and the markup-pricing equation (1) implies that the producer price (the price charged by the manufacturer before taxes) of every variety equals 1. We conclude that producer prices equal 1 independently of tariff rates. Given that skilled labor is used only in the industry supplying the differentiated product and that $f$ units are needed for every brand, the endowment of skilled labor $H^k$ determines the number of brands $n^k (= H^k/f)$ independently of the tariff rate. The reward to skilled workers can be derived from the zero-profit condition, equation (2), which reads $w^k_f = (1 - \alpha)x^k/f$. Hence, the reward to skilled labor increases with output per brand.

In order to see how tariffs affect output per brand, we need to specify preferences over the homogeneous and differentiated products. Take the preference structure

$$U = Y + AD^{\epsilon - 1/\epsilon}, \quad \epsilon > 1,$$

where $Y$ is consumption of the homogeneous good, $A = \epsilon/\epsilon - 1$, and $D = \left[\int d\omega d\omega \omega^{1/\alpha}\right]^{1/\alpha}$, which is a consumption index of differentiated products in which $d\omega$ stands for consumption of variety $\omega$ and the elasticity of substitut-
tion is \( \sigma = 1/(1 - \alpha) > 1 \). This preference structure yields the aggregate demand function

\[
D = P^{-\epsilon},
\]

(6)

and the brand-specific demand function

\[
d_\omega = (p_\omega/P)^{-\sigma}D,
\]

(7)

where \( P \) represents the price index of \( D \):

\[
P = \left[ \int p_\omega^{1-\sigma} d\omega \right]^{1/(1-\sigma)}.
\]

(8)

We can use equations (6) to (8) to derive the demand for a brand manufactured in, say, country 1, in order to examine how this demand changes when country 1 imposes a small tariff on brands imported from country 2. Country 2's demand for country 1's brands does not change when country 1 imposes a tariff, because the tariff does not change consumer prices or the variety choice in country 2. Therefore, country 1's output per brand increases if and only if its own demand for domestic products increases. It is evident from equations (6) to (8), however, that by raising the price index \( P \), a tariff generates two opposing effects. On the one hand, given aggregate demand \( D \), it shifts demand from imported to domestic brands [see equation (7)]. This intra-industry effect tends to increase demand per domestic brand. On the other hand, it reduces aggregate demand for differentiated products [see equation (6)] as a result of substitution toward good \( Y \). This intersectoral effect reduces demand for every brand. The net effect depends on the relative magnitudes of these opposing influences. It is straightforward to see that the intra-industry effect dominates if \( \sigma > \epsilon \) but the intersectoral effect dominates if \( \sigma < \epsilon \). In the former case, output per brand and welfare increase; in the latter, output per brand and welfare decline.\(^1\)

**Variety Choice**

The last tariff issue that I wish to discuss concerns the response of variety choice. As in previous examples, I use a framework that isolates the problem at hand. Venables (1987) provided a suitable model. It is a model in which producer prices and output per brand do not respond to tariffs, so that the entire adjustment takes place via changes in the number of brands. The model is essentially the same as in the previous subsection, except that no skilled labor exists, so that the fixed-cost component is generated by the requirement of \( f \) units of unskilled labor per brand rather than \( f \) units of

\(^1\) This shows that the impression given in Flam and Helpman (1987) and Helpman and Krugman (1989, Chap. 7) that tariffs are useful in the presence of product differentiation is misleading, although it may well be the case that normally \( \sigma > \epsilon \).
skilled labor per brand. In this case, producer prices equal 1 and do not change with ad valorem tariffs on differentiated products (see above). Here, however, output per brand is independent of the tariff rate because the zero-profit condition implies \( x^k = x = f/(1 - \alpha) \), \( k = 1,2 \). Hence, tariffs can affect only the number of brands.

Owing to the model's high degree of linearity, it is necessary to introduce some asymmetry in the demand structure in order to preclude complete specialization in differentiated products. Transport costs of the "melting-iceberg" type will do for this purpose. Let \( \phi > 1 \) represent 1 plus the proportion of the product melted away in transport, and let \( T \) equal 1 plus the tariff rate imposed by country 1 on imports of country 2's brands. Clearing of the product market requires \( x = d^1 + d^2 \) for every brand, while from equations (6) and (7) we have \( d^k = (p^k_\phi - \alpha)(p^k)^\alpha - \epsilon, \ k = 1,2 \). Hence, taking account of transport, tariffs, and the fact that producer prices equal 1, the clearing conditions in countries 1 and 2, respectively, require

\[
\begin{align*}
x &= (P^1)^\alpha - \epsilon + (P^2)^\alpha - \epsilon/\phi^\alpha - 1, \\
x &= (P^1)^\alpha - \epsilon/\phi^\alpha - 1T^\alpha + (P^2)^\alpha - \epsilon.
\end{align*}
\]

It is straightforward to see from this system that the tariff raises \((P^2)^\alpha - \epsilon\) and reduces \((P^1)^\alpha - \epsilon\), as demonstrated in Figure 1. Condition (9a) holds along curve \( x_1x_1 \), while condition (9b) holds along curve \( x_2x_2 \). An increase in country 1's tariff shifts the latter curve upward, as indicated by the broken line. Equilibrium shifts from points 1 to 2.

How the tariff affects the consumer price indexes depends on whether \( \sigma \) exceeds \( \epsilon \) or vice versa. In the former case (in which the intra-industry substitution effect is stronger than the intersectoral substitution effect), the price index of country 2 rises while the price index of country 1 declines. This is an interesting case in which a tariff reduces consumer prices and thereby raises welfare in the tariff-imposing country (there is also a positive revenue effect on welfare). In the latter case, the tariff raises the price index in country 1 and reduces it in country 2.

Since producer prices are constant, the price indexes adjust via the composition of available products \( n^k \). From equation (8) we obtain

\[
\begin{align*}
P^1 &= [n^1 + n^2(T\phi)^1 - \sigma]^{1/(1 - \sigma)}, \\
P^2 &= [n^1\phi^1 - \sigma + n^2]^{1/(1 - \sigma)},
\end{align*}
\]

which imply that the number of products manufactured by a country declines when its price index increases or the price index of its trading partner declines. This point is apparent from Figure 2, in which condition (10a) holds along curve \( P_1P_1 \) and condition (10b) holds along curve \( P_2P_2 \). Equilibrium obtains at point 1. An increase in \( P_i \) shifts the \( P_iP_i \) curve toward
the origin, thereby reducing the equilibrium $n_i$, and raising the equilibrium $n_j$, $j \neq i$.

In addition, the direct effect of the tariff in condition (10a) increases the number of products manufactured in country 1 and reduces the number manufactured in country 2. Hence, when $\sigma > \epsilon$, the tariff reduces the price index in 1 and raises it in 2; the direct as well as the indirect effects lead to an expansion of variety in 1 and to a contraction of variety in 2. When $\sigma < \epsilon$, by contrast, the indirect effects work against the direct effect, but when $\epsilon$ is sufficiently close to $\sigma$, the indirect effect via the price indexes dominates (because the response of the price indexes is very large) and the tariff-imposing country ends up manufacturing fewer brands. The latter point was made by Markusen (1988) in a model with differentiated intermediate inputs. We therefore conclude that a tariff may increase or reduce variety in the protecting country, depending on structure.\(^2\)

\(^2\) The possibility of a contraction of the supply of variety in the tariff-imposing country was disregarded by Venables (1987) and by Helpman and Krugman (1989, Chap. 7), who assumed $\sigma > \epsilon$. 
FIGURE 2
CONSUMER PRICE INDEXES AND VARIETY
MULTINATIONAL CORPORATIONS

To complete the first part of this survey, I return to a structural issue—the role of multinational corporations. Recall the construction of the integrated equilibrium. If trade reproduces the essential features of this equilibrium in the way suggested above, there is of course no room for multinationals. But if the world's structure does not permit factor-price equalization, differences in factor rewards may induce companies to reallocate their activities geographically in order to save costs, even if this requires that they place activities in different countries. This is an important, but not the only, reason for the formation of multinationals (see Caves, 1982). It has interesting implications, however.

The cost-saving motive for the formation of multinationals in industries supplying differentiated products was explored in Helpman (1984, 1985). My exposition follows Helpman and Krugman (1985, Chap. 11). Consider the integrated equilibrium of Chapter 2 and suppose that there are two sectors: sector 1, which manufactures differentiated products with brand-specific economies of scale, and sector 2, which produces a homogeneous product with constant returns to scale. Now suppose that the production of a brand requires headquarters services that are internally supplied by the firm, in addition to the inputs that are directly employed in manufacturing. We can then decompose a firm's employment vector $a_1x_1$ into direct employment in manufacturing $a_{1M}x_1$ and employment in the provision of headquarters services $a_{1Z}x_1$, where $a_{hZ} = a_{hZ}a_{Z1}$ for every input $h$. Here $a_{hZ}$ is the input of $h$ per unit of output of headquarters services $Z$, and $a_{Z1}$ is the input of headquarters services per unit of $X_1$. In this case, the factor-market-clearing condition, equation (3), which reads $V = a_1X_1 + a_2X_2$, can be rewritten as

$$V = a_{1M}X_1 + a_2Z + a_2X_2,$$

where $a_2$ is the input vector per unit of headquarters services and $Z$ ($= a_{Z1}X_1$) is the aggregate output of headquarters services.

Once it is understood that manufacturing facilities can be separated from headquarters and placed in a different country, it becomes evident that prevailing differences in factor rewards induce companies to take advantage of this possibility whenever the two activities need different factor proportions in order to minimize costs. Suppose, for example, that the available inputs are skilled and unskilled labor and the headquarters services are skilled-labor-intensive relative to direct manufacturing. Then companies will desire
to locate headquarters in the country with the relatively cheap skilled labor and manufacturing facilities in countries with the relatively cheap unskilled labor.

Figure 3 is the familiar box diagram showing the allocation of skilled labor $H$ and unskilled labor $L$ in a two-country world. The origin of country 1 is $O^1$, and the origin of country 2 is $O^2$. The vector $O^1Q$ represents total employment of sector 1 in the integrated equilibrium, while $QO^2$ represents total employment in sector 2. Sector 1 is relatively skill-intensive. The parallelogram $O^1QO^2Q'$ represents the factor-price-equalization set of endowments when intra-firm activities cannot be separated and there is a continuum of brands.

**FIGURE 3**

**MULTINATIONAL CORPORATIONS**

Now suppose that manufacturing facilities can be separated from headquarters. We decompose the vector $O^1Q$ into $O^1D$ and $DQ$, where the former is employment in headquarters and the latter employment in direct manufacturing. Headquarters are taken to be more skill-intensive than direct manufacturing. Hence, if the endowment lies in $O^1DQ$, there is no factor-price equalization when headquarters and manufacturing facilities have to be located in the same country, but factor-price equalization obtains when firms from country 1 maintain their headquarters in country 1 but
locate manufacturing in country 2. The possibility of going multinational enlarges the factor-price equalization set.

At world structures in $O^1DQ$ but close to $O^1Q$, the intersectoral pattern of trade resembles the pattern predicted by the Heckscher-Ohlin model. Country 1 imports the homogeneous product relatively intensive in unskilled labor and is a net exporter of differentiated products. There is also intra-industry trade. In addition, however, there is intra-firm trade consisting of exports of headquarters services by multinationals headquartered in country 1 to their subsidiaries in country 2. Helpman (1985) and Helpman and Krugman (1985, Chap. 12) showed how to extend this analysis to include intra-firm trade in intermediate inputs. In any case, the degree of multinationality and the volume of intra-firm trade increase with the difference in the ratios of factor endowments (the further away the location of the endowment point from $O^1Q$). By the same token, the larger the difference in factor proportions, the smaller are country 1's net exports of differentiated products, because more products are produced in country 2. At points close to $O^1D$, country 1 imports differentiated products because its resources are mainly employed in the production of headquarters services.

One can in fact show that in $O^1DQ$ a larger difference in factor proportions raises the share of intra-firm trade, and for points close to $O^1Q$ it also raises the share of intra-industry trade (holding constant relative country size). The latter relationship is just the opposite of what happens in the absence of multinationals. It stems from the fact that more brands are moved to country 2 rather than to country 1 when the difference in factor proportions widens. When country 1 is a net importer of differentiated products, larger differences in factor proportions reduce the share of intra-industry trade. Hence, the emergence of multinational corporations introduces a nonmonotonic link between differences in factor proportions and the share of intra-industry trade. Despite these elaborate links and a changing pattern of commodity trade, the model preserves a fundamental property of the factor-proportions theory. Multinationals notwithstanding, every brand is produced with the same vector of inputs per unit of output and both countries consume the same composition of products. Therefore, the factor content of net trade flows obeys Vanek's chain rule: the country relatively rich in skilled labor is a net exporter of skilled-labor services and a net importer of unskilled-labor services. Naturally, this requires counting the factor content of intra-firm trade in headquarters services as part of the general international flow of factor services.
Thus far, my discussion has been restricted to static issues. In recent years, however, the theory of trade in differentiated products has been extended to dynamic environments. The need for such extensions is obvious. Many concerns cannot be satisfactorily addressed with static models, such as the evolution of trade over time or the effects of trade policy on long-run growth. In order to demonstrate what can be achieved with dynamic models of product differentiation, I describe a line of work that Gene Grossman and I have explored.

In static models, fixed costs are often interpreted as an outlay for product development. It is clear upon reflection, however, that the development of a brand should be studied in a dynamic context, in which an entrepreneur incurs development costs first and collects operating profits later. As long as the expected present value of operating profits covers the brand-development costs, there is an incentive to invest in R&D. Investment in R&D leads to the accumulation of products and thereby to growth.

Consider an environment in which the development of a new brand entitles the entrepreneur to indefinite monopoly power (I deal with limited monopoly power later). There is free entry into product development, and active R&D takes place over an interval of time. Then at each time \( t \) in this interval, product-development costs \( c_n[w(t)] \) must equal the present value of future operating profits \( \pi(\tau) \):

\[
c_n[w(t)] = \int_t^\infty \exp[R(\tau) - R(t)]\pi(\tau)d\tau ,
\]

where \( w(t) \) is the vector of input prices and \( R(t) \) is the discount factor from \( t \) to 0 (see Judd, 1985). This condition replaces the zero-profit requirement (2), and it implies the fundamental asset-pricing equation, or no-arbitrage condition,

\[
\pi/c_n + \dot{c}_n/c_n = \dot{R} ,
\]

which states that the instantaneous profit rate \( \pi/c_n \) plus the capital gain on the value of the firm \( \dot{c}_n/c_n \) equals the instantaneous interest rate. The firm derives its value from the claim to future profits. This value equals the present value of the profit stream, on the one hand, and the cost of developing a new brand that will provide the same profit stream, on the other.

Grossman and Helpman (1989d) embedded this approach in a dynamic trade model that abstracts from other sources of growth, such as capital accumulation. In order to see its implication, it is best to begin with the
integrated equilibrium. Production requires skilled and unskilled labor with fixed input vectors per unit of output: \( a_n \) per unit in R&D (which leads to an incremental change in the measure of available brands), \( a_1 \) per unit in the manufacture of a differentiated product, and \( a_2 \) per unit in the manufacture of a homogeneous product. Dynamic economies of scale derive from the fact that after the fixed product-development costs \( w(t) \cdot a_n \) are incurred at time \( t \), all future variable costs per unit of output equal \( w(t) \cdot a_1 \). Using the R&D costs as numeraire and assuming symmetric CES preferences for differentiated products, with an elasticity of substitution \( 1/(1 - \alpha) \), the pricing equations are

\[
\begin{align*}
1 & = w \cdot a_n , \\
\alpha p_1 & = w \cdot a_1 , \\
p_2 & = w \cdot a_2 ,
\end{align*}
\]

and the factor-market-clearing condition becomes

\[
\begin{bmatrix}
L \\
H
\end{bmatrix} = a_n \dot{n} + a_1 X_1 + a_2 X_2 ,
\]

(14)

where \( L \) is unskilled labor, \( H \) is skilled labor (human capital), and \( \dot{n} \) is the measure (number) of newly developed products (i.e., the time derivative of the number of available products \( n \)).

Suppose, in addition, that consumers allocate a proportion \( \gamma_i \) of aggregate spending to sector \( i \)'s goods. Then

\[
p_i X_i = \gamma_i E , \quad i=1,2,
\]

(15)

where \( E \) is aggregate spending. Now equations (13) to (15) can be used to solve for prices, factor rewards, product development, and output levels as functions of consumer spending \( E \). In particular,

\[
\dot{n} = \nu(E) .
\]

(16)

As consumers increase their spending, more resources are drawn into production of consumer goods, thereby leaving fewer resources available for R&D purposes. Consequently, product development slows down.

To complete the analysis, we need to specify savings. If all consumers are alike and allocate spending over time according to a time-additively-separable preference function that is logarithmic in spending, the rate of growth of spending obeys \( \dot{E}/E = \bar{R} - \rho \), where \( \rho \) is the subjective discount rate. The choice of numeraire implies constant R&D costs, so that the asset-pricing equation (13a) yields \( \bar{R} = \pi \) (i.e., the interest rate equals the profit rate). Furthermore, equation (13b) implies that profits per product are a proportion \( (1 - \alpha) \) of expenditure on the product, so that using equation (15) profits per brand are \( \pi = (1 - \alpha) \gamma_i E/n \). Therefore,
\[ \dot{E}/E = (1 - \alpha)\gamma, E/n - \rho . \] 

(17)

The autonomous differential-equation system (16) and (17) can be solved for every initial value of the number of products and consumer spending. For every initial small value of the number of products, however, there is only one initial spending level for which the solution satisfies the consumer's transversality condition. Hence, there exists a unique perfect-foresight equilibrium trajectory. This trajectory converges to a steady state in which product development ceases. On the way to the steady state, the number of products and consumer spending rise over time.

We can use Figure 4, which is an elaboration of Figure 3, to describe the evolution of sectoral employment levels. Let product development be the activity most intensive in skilled labor, followed by the manufacture of differentiated products, leaving the manufacture of the homogeneous product as most intensive in unskilled labor. Given the number of products, let \( O_1^D \) be employment in R&D, \( DQ \) be employment in the manufacture of differentiated products, and \( QO^2 \) be employment in the production of homogeneous products. As time goes by, employment in R&D declines, to \( O_1^D \), for example, employment in the manufacture of differentiated products rises to \( D_1 Q_1 \), and employment in the manufacture of \( X_2 \) declines to \( Q_2 O^2 \). In the limit, when product development comes to an end, \( X_1 \) employs \( O^1 Q_\infty \) and \( X_2 \) employs \( Q_\infty O^2 \).
What happens when we decompose this world into two countries? First take a structure in \(O^1Q_\infty O^2\), where country 1 is relatively rich in skilled labor. Grossman and Helpman (1989d) have shown that for every allocation of factor endowments to countries in this set there is a unique, constant product ratio \(n_1(t)/n_2(t)\) that is consistent with a starting situation in which brands do not exist. If the world begins with this ratio, countries will invest in R&D in proportion to the available number of products and therefore the number of products will grow at the same rate in both countries. Consequently, the initial product ratio will be preserved forever. In addition:

1. There will be factor-price equalization at each point in time, although the wages of skilled workers will decline over time while the wages of unskilled workers will rise over time.
2. Country 1 will be a net exporter of differentiated products and a net importer of the homogeneous product.
3. World trade will grow faster than GNP.

Naturally, there will also be intra-industry trade. This trajectory reproduces the essential features of the integrated equilibrium, and at no time is there any incentive to go multinational.

The nature of the equilibrium trajectory changes when the endowment is in \(O^1Q_\infty\). Initially, it looks the same as that described above. But total employment in the differentiated-product sector—in R&D plus manufacturing—becomes more intensive in unskilled labor over time as a result of the relative decline of product development, and the joint employment vector \(O^1Q\) rotates clockwise, to \(O^1Q_1\) initially and to \(O^1Q_\infty\) eventually. Therefore, there comes a time when this joint employment vector reaches the endowment point. From then on, factor-price equalization cannot be preserved without companies' going multinational. It is quite clear from the discussion in the previous chapter that at this stage companies based in country 1 have an incentive to carry out at home the activities most intensive in skilled labor and to place abroad the activities most intensive in unskilled labor. Now suppose that a production vector, such as \(DQ\), consists of employment in headquarters services and direct manufacturing. Specifically, let headquarters services require only skilled labor and direct manufacturing require only unskilled labor. No matter where the endowment point is in \(O^1Q_\infty\), the separation of direct manufacturing from headquarters enables multinationals based in country 1 to perform product development at home, locate headquarters services at home, and place the manufacture of differentiated products in country 2. Indeed, in this case the emergence of multinational corporations recovers factor-price equalization and the essential features of the integrated equilibrium. As time goes by, more
resources are employed by subsidiaries of multinational corporations. The steady state, in which no further product development takes place, is very similar to the static equilibrium described in the previous chapter. (But if headquarters services also use unskilled labor, the set of endowments for which multinationals reproduce this equilibrium is somewhat smaller.)
In Chapter 6 I outlined a dynamic model of product innovation in which growth peters out over time and the world reaches a static steady state. This happens for the same reason that growth peters out in a neoclassical model of capital accumulation. In order to see the analogy, recall the dynamic spending equation $E = R - \rho$. In a neoclassical economy, the interest rate equals the marginal product of capital. If the labor force does not change, capital accumulation depresses the interest rate. (It is easy to extend the argument to an economy with a growing labor force). Therefore, in an economy that starts with a small capital stock, the interest rate is higher than the subjective discount rate $\rho$, spending rises over time, and the interest rate declines as more capital is accumulated. When the interest rate equals the subjective discount rate, capital accumulation ceases, as does the growth in spending. It is therefore clear that sustained long-run growth can prevail in this type of an environment only if the marginal product of capital does not decline to the subjective discount rate. This is achieved, for example, with economies of scale in the use of capital that are external to the firm but internal to the economy (see Romer, 1986).

In the Grossman-Helpman model, no capital accumulation takes place, and the interest rate is determined by the profit rate. The profit rate exceeds the subjective discount rate initially, leading to rising spending and growth in the number of products. As more brands compete for consumer spending, profits per brand decline and the interest rate follows suit. Growth of variety ceases when the profit rate, and with it the interest rate, reaches the subjective discount rate. It is clear from this description that the number of products plays the role of a capital stock; investment takes the form of R&D, and the accumulation of variety depresses the rate of return on investment. Since the lower bound on the profit rate is zero, the rate of return on investment must decline to the point at which growth peters out.

In the rest of this chapter I show how sustained long-run growth can prevail in the presence of learning by doing in product development. This insight is then used to explain endogenous rates of innovation and imitation in a North-South model of the product cycle and to discuss the effects of trade and industrial policy on long-run growth.

Endogenous Growth

Arrow's (1962) celebrated concept of learning by doing can be invoked to posit the existence of external economies in the growth of products, which
can prevent the zero-growth outcome. This idea was formalized by Romer (1988). The following exposition relies on a simplified version devised by Gene Grossman and me. Suppose that one type of labor $L$ serves as an input in R&D and in the manufacture of differentiated products for which there is a CES preference function. There are no other goods. Manufacturing requires $a_{LX}$ units of labor per unit of output, while product development requires $a_{LX}/K$ units of labor per product, where $K$ represents the stock of knowledge capital in product development. The larger this stock the lower is labor input per unit of R&D output. Suppose that the stock of knowledge reflects learning by doing according to $K = \dot{n}$. Then, by a proper choice of units, $K = n$ and the pricing equations (13a) and (13b) become

\[
\begin{align*}
1 &= wa_{LX}/n, \\
\alpha p &= wa_{LX}.
\end{align*}
\] (18a)

(18b)

Pricing and the choice of numeraire imply that wages and prices grow at the rate of product growth. Therefore, from the no-arbitrage condition, equation (12), the interest rate equals the profit rate.

Now the factor-market clearing condition analogous to equation (14) reads $a_{LX}n/n + a_{LX}X = L$. But since consumer spending $E$ equals $pX$, we can use equation (18) to write it as

\[
\eta/n = L/a_{LX} - \alpha \eta,
\] (19)

where $\eta = E/n$ represents consumer spending per product.

Profits per product equal $(1 - \alpha)\eta$, which also equals the interest rate $\dot{R}$. Using this result together with the growth-of-spending equation $\dot{E}/E = \dot{R} - \rho$ and with equation (19), we obtain

\[
\dot{\eta}/\eta = \eta - \rho - L/a_{LX}.
\] (20)

This differential equation in spending per product has only one solution that satisfies the consumers' transversality condition: the stationary solution. Therefore, the economy jumps immediately to a steady state in which spending per product remains constant. The implied growth rate of products $g = \dot{n}/n$ is obtained from equations (19) and (20):

\[
g = (1 - \alpha)L/a_{LX} - \alpha \rho.
\] (21)

Hence, the economy grows faster the larger the effective labor force in terms of R&D $(L/a_{LX})$, the lower the subjective discount rate, and the higher the degree of monopoly power (as measured by the markup of price over marginal cost $R = 1/\alpha$).

As in the North-South-trade example in Chapter 3, the welfare of a typical worker at a point in time can be measured by $(w/p)n^\alpha - \alpha \dot{p}$ or its logarithm. Since the wage/price ratio is constant, the worker's temporal well being rises over time at the rate $g(1 - \alpha)/\alpha$. In this case, a worker is better off
living in a large economy that starts with the same number of products as a smaller economy, because his initial level of well-being is the same in both countries but rises faster in the larger country.

The North-South Product Cycle

Vernon suggested in 1966 that North-South trade follows a product cycle in which the North develops new products that it manufactures for a while, but the lower-cost South takes over production after the product is sufficiently standardized that the South can learn to make it. Hence, high-tech products (electronics, at that time) are exported first by the North and later on by the South. Krugman (1979a) formalized this idea using a one-sector model with product differentiation in the following way:

Only the North knows how to develop brands. The rate of growth in the number of brands $g$ is exogenous. The South knows only how to imitate products. In particular, if $n_s$ represents the number (measure) of brands that the South knows how to produce and $n_N (= n - n_s)$ represents the number of brands in which the North maintains monopoly power, the rate of imitation equals $\dot{n}_s/n_N$. Krugman assumed an exogenous rate of imitation $\mu$. All this implies a differential equation for the share of products $\sigma_s (\equiv n_s/n)$ that the South knows how to produce: $\dot{\sigma}_s/\sigma_s = \mu(1 - \sigma_s)/\sigma_s - g$.

This is a stable differential equation that converges to a steady state in which the South's share in available variety equals $\mu/(g + \mu)$; the larger the rate of imitation and the smaller the rate of innovation, the larger is the South's share. Given the exogenous rates of innovation and imitation, the evolution of the number of products and their composition is not driven by economic forces; rather, it is mechanical. This contrasts with the previous section in which economic considerations were important determinants of the rate of innovation. Economic considerations should likewise be important determinants of the rate of imitation, because imitation also requires resources and thus involves an economic cost-benefit calculation.

Following Grossman and Helpman (1989a), suppose that innovation takes place in the North according to the considerations described in the previous section (the simple one-sector, one-factor model), except that now a Northern entrepreneur knows that he is not assured of indefinite monopoly power because of Southern imitation. Given an instantaneous rate of imitation $\mu = \dot{n}_s/n_N$, every Northern brand has the same chance of being imitated. In this case, the rate of imitation equals the time $\tau$ hazard rate of the cumulative distribution function $F(t,\tau)$, which describes the probability that a product developed at time $t$ will be imitated by the South before $\tau \geq t$. Assuming that Northern entrepreneurs maximize the expected present value of their profits using this distribution function, it can be shown that the no-arbitrage condition, equation (12), should be replaced by
\[ \frac{\pi}{c_n} + \frac{\dot{c}_n}{c_n} = \dot{R} + \mu . \]  

The instantaneous profit rate plus the capital gain equal the interest rate plus a risk premium that represents the conditional density of losing monopoly power (a capital loss). The factor-market-clearing condition and the pricing equations remain essentially the same.

Grossman and Helpman have postulated that it takes resources to imitate a product. Specifically, a Southern entrepreneur needs \( a_{LI} \) units of labor per brand for imitation. Having imitated a variety, she needs \( a_{LX} \) units of labor per unit of output in manufacturing (just like the North). Imitation takes place only if the present value of profits covers imitation costs. For an imitator, however, the profit calculation is more involved. If she did not face competition from the original Northern innovator, she would mark up price above marginal costs in the usual way. When the resulting price falls short of Northern marginal manufacturing costs, she can still charge this price without being threatened by the Northern producer. This happens when the South's wage rate is lower than the proportion \( \alpha \) of the North's wage rate and is termed the "wide gap" case (the gap in relative wages is wide). Otherwise, the Southern imitator charges a price that equals the North's marginal manufacturing costs. Naturally, the imitator would lose money in either case if the wage rate were lower in the North, so that active imitation requires a lower wage rate in the South, and this is assumed hereafter. Free entry into imitation implies a no-arbitrage condition such as equation (12).

In addition, assume that the stock of knowledge capital in imitation equals \( n_s \) (this can be extended). Therefore, labor-market clearing in the South implies

\[ a_{LI} \dot{n}_s / n_s + a_{LX} X_s = L_s . \]

Now assume the wide-gap case, so that the South's pricing equations are similar to those of the North. Together with the market-clearing and no-arbitrage conditions and the growth-of-spending equation (in which the subjective discount rate is the same in both countries), the pricing equations imply a steady-state growth equation that is analogous to equation (21):

\[ g = \frac{(1 - \alpha) L_s / a_{LI} - \alpha \rho}{\alpha g + \mu} . \]  

A similar procedure for the North, using equation (22), yields an equilibrium steady-state relationship between the rate of innovation and the rate of imitation:

\[ (1 - \alpha)(L_n / a_{LN} - g)(g + \mu) / \alpha g = g + \mu + \rho . \]  

The left-hand side represents the profit rate, while the right-hand side represents the interest rate plus the risk premium. The right-hand side
increases in $g$, while the left-hand side declines in $g$. Therefore, an increase in the rate of innovation reduces profitability relative to the cost of capital. Alternatively, an increase in $\mu$ raises the right-hand side but raises the left-hand side even more. Therefore, an increase in the rate of imitation increases the profitability of innovation relative to the capital cost. This explains the upward slope of curve $NN$ in Figure 5 along which equation (24) holds. As equation (23) holds along $SS$, the equilibrium levels of innovation and imitation are given by the intersection point 1.

**FIGURE 5**

**NORTH-SOUTH PRODUCT CYCLE**

Several implications of this model are worthy of note. First, observe that if innovation in the South requires more resources than imitation, which is reasonable, then trade with the North speeds up long-run growth in the South. This can be seen from equation (23). Without trade, the growth equation is the same [see (21) for a closed economy], except that $a_{Li}$ is replaced by a larger coefficient. Second, trade with the South speeds up long-run growth in the North. This is shown in Figure 5 by the fact that the
vertical intercept of \( NN \) identifies the autarky growth rate, so that both countries grow faster by trading with each other. Third, it is clear from Figure 5 that a larger South raises both the rate of innovation and the rate of imitation. A larger North, by contrast, does not affect the rate of innovation but reduces the rate of imitation. (The rate of innovation increases if knowledge capital in imitation also depends on the number of unimitated products.) Lower rates of imitation are associated with longer average time periods during which Northern entrepreneurs command monopoly power. Fourth, it can be shown that the larger a country, the larger is its relative wage rate. This stands in sharp contrast with the results obtained in both the static North-South trade model discussed in Chapter 3 and Krugman’s model with exogenous rates of innovation and imitation. Indeed, it is precisely the endogeneity of these rates that drives the result, because it works through economies of scale to offset the static relative-scarcity effect.

**Policy**

A central feature of the growth process I have discussed is that the long-run growth rate depends on structural features such as the resource bases of the trading partners, the degree of monopoly power, and productivity levels. These factors affect the growth rate because they affect equilibrium employment in product innovation and its contribution to the inflow of new goods. Thus, if \( L_n \) represents employment in R&D, the long-run growth rate \( g \) equals \( L_n/a_{Ln} \). In the example in the section on endogenous growth, where we considered a single country, \( L_n \) stands for that country’s employment in R&D. In the product cycle of the preceding section, employment in innovation in the North is the relevant variable.

In fact, it is instructive for the purpose at hand to reinterpret the single-country example in a multi-country context in order to identify the role of country-specific innovation. Suppose there are two countries with the characteristics specified above and that they are identical in all respects. Now consider the case in which the R&D experience of one country contributes the same amount of knowledge capital to both countries. In other words, knowledge is disseminated across national boundaries just as it is within national boundaries. We can therefore aggregate the two economies into one double-sized economy and apply the one-country analysis. The common long-run growth rate will depend on aggregate world employment in innovation, which will be determined by structural features. Each country will contribute one-half the innovation effort. In more general environments, where countries are not identical, long-run growth depends in a complicated way on the world’s structure. Asymmetries that can play an important role include size, comparative costs, and differential dissemination rates of knowledge at home and abroad (see Grossman and Helpman, 1989a).
Once we recognize that the allocation of resources to R&D determines growth rates, it becomes evident that every policy that induces a long-run resource reallocation affecting employment in R&D also changes long-run growth rates. In a multi-country system, moreover, one country's policies may affect R&D levels at home and abroad in opposite directions and the net effect on growth can be positive or negative (see Grossman and Helpman, 1989a).

As an example, consider the product-cycle model in the previous section. Other things equal, a subsidy to innovation in the North raises profits on newly developed products. This can be represented by a parameter that multiplies the left-hand side of equation (24). Its value equals 1 in the absence of a subsidy and exceeds 1 in the presence of an R&D subsidy. In terms of Figure 5, the subsidy shifts the NN curve upward. In the wide-gap case, the equilibrium point shifts to the left on the SS curve, because the latter does not shift with the subsidy. Hence, the innovation subsidy slows down imitation but does not change the rate of innovation. In the narrow-gap case, Grossman and Helpman (1989b) show that an innovation subsidy speeds up both imitation and innovation. Coming back to the wide-gap case, an imitation subsidy shifts the SS curve upward and does not affect the NN curve. Therefore, it speeds up imitation as well as innovation.

Are growth-promoting policies desirable? It is easiest to answer this question for the simple one-country model, where growth is definitely undersupplied. The bias stems from the fact that R&D today generates a positive externality for R&D in the future, because it contributes to knowledge capital that reduces future innovation costs. (An innovator similarly fails to take into account his contribution to consumer surplus and the reduction of profits on other brands, but these effects cancel out.) For this reason, some degree of growth promotion by means of R&D subsidies is desirable. However, excessive subsidization of product development speeds up growth to the point at which welfare declines.

In order to see this more precisely, we can follow Grossman and Helpman (1989c). Aggregate welfare at a point in time is measured by the logarithm of \((E/p)n^{\delta - \alpha/\alpha}\) similar to the North-South trade model of Chapter 3. Taking account of the fact that \(E, p,\) and \(n\) all grow at the rate \(g\), the discounting of this utility flow by means of the subjective discount rate yields our welfare measure \(U\), which obeys

\[\rho U = \text{constant} + \log \eta + g(1 - \alpha)/\alpha \rho,\]

where the constant depends on the initial number of products. This function represents an induced preference ordering over expenditure per product \(\eta\) and the growth rate \(g\). We calculate the tradeoff between them by maximizing \(\rho U\) subject to the resource constraint (19): \(g + \alpha \eta = L/a_{L_n}\). The
optimal growth rate is larger than the equilibrium growth rate (21). Hence R&D subsidies that bring the growth rate closer to the optimal level raise welfare, but further increases in the subsidy rate reduce welfare.

It should be clear by now that commercial policy can also affect long-run growth rates (see Grossman and Helpman, 1989a,c). If, for example, trade policy succeeds in diverting resources toward product innovation, it accelerates growth. But even in cases where the free-trade growth rate falls short of the optimal level and trade policy accelerates growth, trade policy may nevertheless be harmful (see Grossman and Helpman, 1989c). Monopolistic competition per se introduces a distortion that can be aggravated by a growth-enhancing trade policy. In addition, in the presence of rent seeking, growth is slower under quotas than under tariffs because quotas divert resources to rent seeking, thereby reducing employment in R&D (see Grossman and Helpman, 1989c). This effect is particularly strong when rent seeking uses entrepreneurial skills that are useful in product development. These examples highlight the role of policy in a dynamic context.
8 CONCLUDING COMMENTS

The new static model represents a natural progression from, and enrichment of, the factor-proportions theory. The new dynamic models extend the static approach in a way that enables us to address important new issues. Both incorporate elements of industrial organization and bring the theory closer to the "real world." It is now recognized that the "simplifying" assumption of perfect competition is sometimes very costly in terms of relevance to the world around us and our ability to explain observed phenomena. This conclusion applies with equal force to structural features and to policy implications, as I have demonstrated in this paper. I have dealt only with monopolistic competition, but the new approach has considered other market structures as well. In short, international trade theory has taken advantage of a new framework that has brought it closer to reality than ever before.
REFERENCES


PUBLICATIONS OF THE
INTERNATIONAL FINANCE SECTION

Notice to Contributors

The International Finance Section publishes papers in four series: ESSAYS IN INTERNATIONAL FINANCE, PRINCETON STUDIES IN INTERNATIONAL FINANCE, and SPECIAL PAPERS IN INTERNATIONAL ECONOMICS, all of which contain new work not published elsewhere, and REPRINTS IN INTERNATIONAL FINANCE, which reproduce journal articles published elsewhere by Princeton faculty associated with the Section. The Section welcomes the submission of manuscripts for publication as ESSAYS, STUDIES, and SPECIAL PAPERS.

ESSAYS are meant to disseminate new views about international financial matters and should be accessible to well-informed nonspecialists as well as to professional economists. Technical terms, tables, and charts should be used sparingly; mathematics should be avoided.

STUDIES and SPECIAL PAPERS are meant for manuscripts too long for journal articles and too short for books. Hence, they may be longer and more technical than ESSAYS. STUDIES are devoted to new research on international finance (with preference given to empirical work). They should be comparable in originality and technical proficiency to papers published in the leading journals. SPECIAL PAPERS are surveys of research on particular topics and should be suitable for use in undergraduate courses. They may be concerned with international trade as well as international finance.

Submit three copies of your manuscript. Retain one for your files. Your manuscript should be typed on one side of 8½ by 11 strong white paper; all materials should be double spaced, including quotations, footnotes, references, and figure legends. For further guidance, write for the Section’s style guide before preparing your manuscript, and follow it carefully.

How to Obtain Publications

The Section’s publications are distributed free of charge to college, university, and public libraries and to nongovernmental, nonprofit research institutions. Eligible institutions should ask to be placed on the Section’s permanent mailing list.

Individuals and institutions that do not qualify for free distribution can receive all publications by paying $30 per calendar year. (Those who subscribe during the year will receive all publications issued earlier that year.)

Publications can be ordered individually. ESSAYS and REPRINTS cost $6.50 each; STUDIES and SPECIAL PAPERS cost $9. Payment must be made when ordering, and $1.25 must be added for postage and handling. (An additional $1.50 should be added for airmail delivery outside the United States, Canada, and Mexico.)

All payments must be made in U.S. dollars. (Subscription fees and charges for single issues will be waived for organizations and individuals in countries whose foreign-exchange regulations prohibit such payments.)

Address manuscripts, correspondence, subscriptions, and orders to:

International Finance Section
Department of Economics, Dickinson Hall
Princeton University
Princeton, New Jersey 08544-1017

Subscribers should notify the Section promptly of any change of address, giving the old address as well as the new.
List of Recent Publications

To obtain a complete list of publications, write the International Finance Section.

ESSAYS IN INTERNATIONAL FINANCE

152. G. K. Helleiner, The IMF and Africa in the 1980s. (July 1983)
163. Arminio Fraga, German Reparations and Brazilian Debt: A Comparative Study. (July 1986)
167. Rainer Stefano Masera, An Increasing Role for the ECU: A Character in Search of a Script. (June 1987)


---

**PRINCETON STUDIES IN INTERNATIONAL FINANCE**

49. Peter Bernholz, *Flexible Exchange Rates in Historical Perspective.* (July 1982)


60. Thorvaldur Gylfason, *Credit Policy and Economic Activity in Developing Countries with IMF Stabilization Programs.* (Aug. 1987)


---

* Out of print. Available on demand in xerographic paperback or library-bound copies from University Microfilms International, Box 1467, Ann Arbor, Michigan 48106, United States, and 30-32 Mortimer St., London, WIN 7RA, England. Paperback reprints are usually $20. Microfilm of all Essays by year is also available from University Microfilms. Photocopied sheets of out-of-print titles are available on demand from the Section at $9 per Essay and $11 per Study or Special Paper, plus $1.25 for postage and handling.


64. Jeffrey A. Frankel, *Obstacles to International Macroeconomic Policy Coordination.* (Dec. 1988)


SPECIAL PAPERS IN INTERNATIONAL ECONOMICS


REPRINTS IN INTERNATIONAL FINANCE


