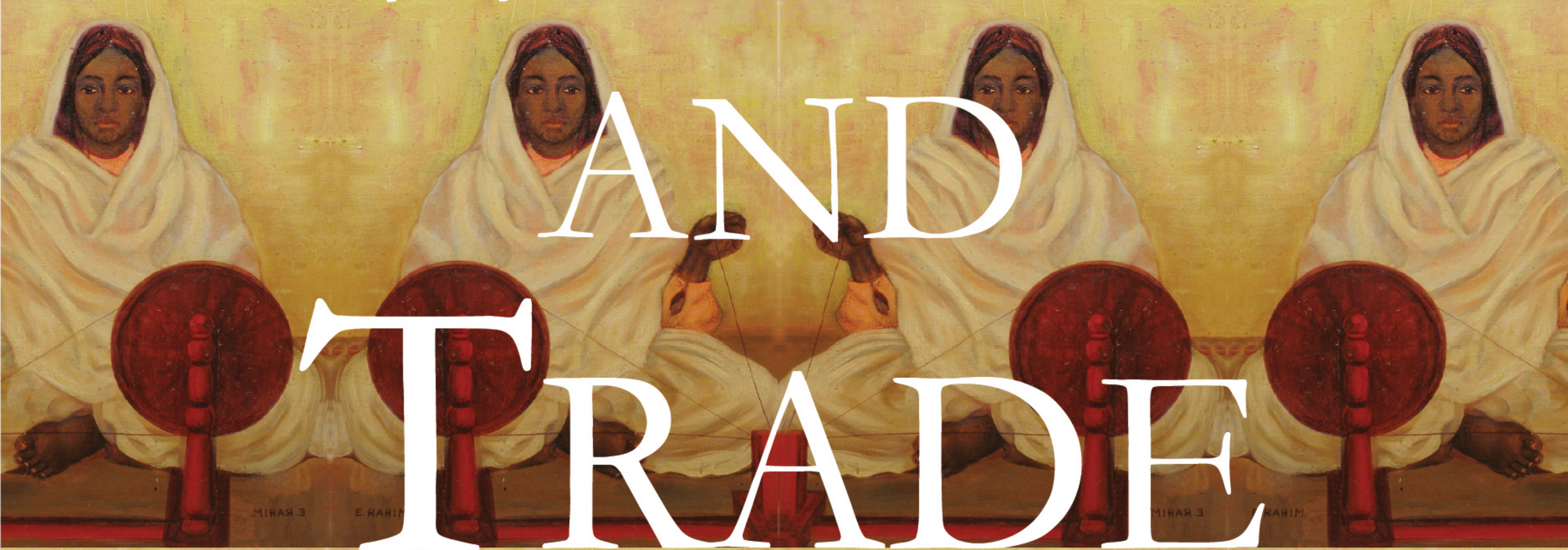
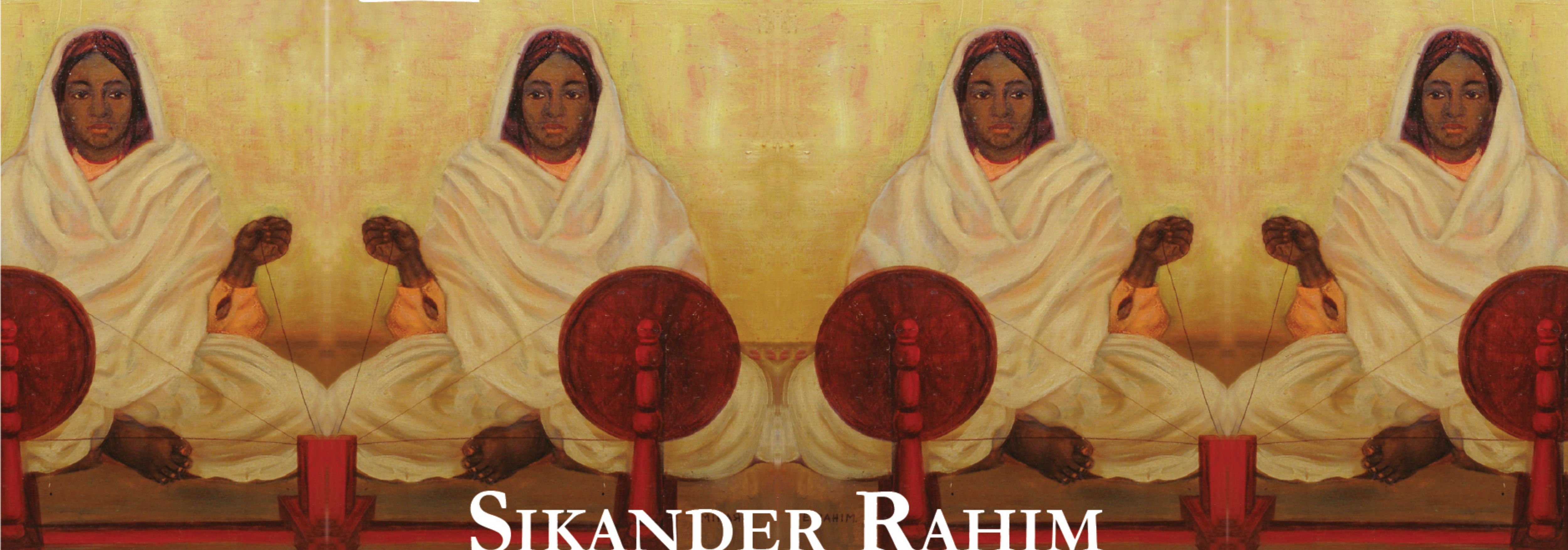




WAGES



AND
TRADE



SIKANDER RAHIM



HOW WAGES AND TECHNICAL PROGRESS
DETERMINE INTERNATIONAL TRADE

WAGES AND TRADE

How Wages and Technical Progress Determine International Trade

Sikander Rahim

“The ideas which are expressed here so laboriously are
extremely simple and should be obvious.”

J. M. Keynes

The General Theory of Employment, Interest and Money

“For between true Science and erroneous Doctrines,
Ignorance is in the middle.”

Thomas Hobbes

Leviathan



Lahore School of Economics

Copyright © 2020 by Sikander Rahim under exclusive license to the Lahore School of Economics for publishing and printing the first edition in both print and electronic form.
All rights reserved

First Edition: August, 2020
ISBN: 978-969-7502-08-0
Price: Rs.2000/-

The views expressed in this document are those of the author and do not necessarily reflect the views of the Lahore School of Economics.

Printed and published by the Lahore School of Economics,
Intersection Main Boulevard, Phase VI, DHA, and Burki Road Lahore
53200, Pakistan
Tel.: +92 42 3656 0936
<http://www.lahoreschoolofeconomics.edu.pk>

Table of Contents

FOREWORD	i
CHAPTER 1	1
Comparative Advantages and Capital Goods	1
1. Introduction	1
2. Trade with Capital Goods	9
3. The Factor Endowments Theory: What is a Factor?.....	19
4. The Factor Endowments Theory: General Equilibrium and Produced Factors.....	26
5. The Factor Endowments Theory: Comparison with Reality	33
CHAPTER 2	49
Differentiated Goods, Brands and Technical Progress	49
1. Differentiated Goods and Brand Names	49
2. Simple Goods, Complex Goods and Technical Progress	65
3. Economic Growth Since World War II	94
CHAPTER 3	125
The Balance of Trade and the Exchange Rate	125
1. National Product, Expenditure and the Trade Balance	125
2. Theories of the Trade Balance and Exchange Rates.....	153
3. International Payments: the Bretton Woods System and After.....	187
4. The Effects of Changes of Exchange Rates on Low Wage Countries	223
CHAPTER 4	229
Doctrine and Consequence: Pakistan's Case	229
1. Historical: First Stages	229
2. Value Added at World Prices	232
3. Globalisation and the Ill-Adapted Economy.....	235
4. Poverty and Pessimism.....	259
APPENDIX	269
Capital, Technical Progress and International Trade	269
1. Capital Goods and Determinacy	270
2. Capital Goods and Technical Progress	289
3. Summing Up.....	310
REFERENCES	313
INDEX	327
ERRATUM	335

FOREWORD

All theories about international trade have had their doubters. Theory has a psychological advantage over doubt; it meets the human desire for explanation. Its advantage is all the greater among the prosperous when the explanation it purports to provide is in harmony with the prevailing beliefs about production and prices. Those who have prospered less are more inclined to doubt. But doubters are at a disadvantage; if they reject the explanation without providing an alternative, they leave the subject of international trade in limbo. Their grounds for doubt, empirical or theoretical, can be sound and it may be correct scientific practice to assert ignorance, but it is hard not to abhor a vacuum and to leave a subject of that importance without a theory. Economists have preferred constructing theories, more than one, to living with doubt and, as though there were no room for doubt, they have propagated their theories as zealously as did physicians once their humours.

Their zeal has been all the more for a second advantage of their theories, that of concluding that free markets work best and, in particular, that free trade benefits all countries. The orthodoxy has changed from time to time, but the conclusion has been modified only slightly. At first there was not much to this, no more than the assertion of Adam Smith and some of his predecessors that nothing was gained from choosing a domestically made good if the same good imported was cheaper. Later, in 1815, David Ricardo provided the first full theory of international trade, by which is meant that it purports to show how the whole economy adjusts to trade. Eli Heckscher and Bertil Ohlin formulated an alternative theory roughly a century later. Both are theories of comparative advantage being full theories and asserting that, given consumer preferences, the production and trade of each country are fully determined in free trade by how countries compare with respect to certain characteristics, these characteristics being in Ricardo's theory the amounts of labour needed per unit of output and in the Heckscher-Ohlin theory the amounts of the various factors with which each country is endowed.

What has changed is that the industrialised countries did not as a rule propound free trade before they had industrialised but do so now. In America in 1791 Alexander Hamilton stated that a prerequisite for industrialisation is protection against the competition of the established industrial economies, for an industry new to a country needs time and experience to become competitive with the established industries of other countries. America remained a protectionist country for the next one and a half centuries and has been advocating trade liberalisation since (with some

exceptions, including agriculture). Hamilton was disputing Smith's implicit assumption that neither present nor future costs are affected by present production and provided the basis of the infant industry argument that Friedrich List took up from his followers and which has been regularly used since to argue for protection. List brought about the customs union of German states, the *Zollverein*, and Germany is now for free trade (except for agriculture). But neither List nor any other likeminded person could provide a full theory. It is practically impossible to do so when costs depend on past production, if only because the reasoning or mathematical formulae are intractable for economic theorists. Much the same follows from disputing the premise of constant returns to scale, for to do so makes costs depend on the amount of production.

Theory may have a psychological advantage, but such evidence as history provides seems mostly to have been in favour of the doubters. There seems to be no instance of an industrial economy of any significance of which the industry was not in some way protected in the early stages – trade barriers, geographical barriers, war and even destruction of the competition. Countries also provided their exporters with other “aids, bounties and premiums”, which Hamilton thought to be the greatest obstacle to a new branch of industry in the other countries.

And yet, since the Second World War the pattern of protection in the world's trade has conformed neither to what would have been expected from theory nor to the doubters' infant industry argument. For the first five decades or more after the War the high wage industrial economies of Western Europe and North America protected their domestic industries against the competition of the exports of what can be called the low wage countries. As they liberalised trade among themselves they cooperated in creating an elaborate system of quotas and tariffs, including the General System of Preferences (GSP), restricting the exports of the low wage countries. Their governments and economists, nonetheless, always insisted on their belief in the benefits from free trade and their protectionism cannot be explained by the infant industry argument or returns to scale, for their industries were long established, experienced and bigger than those of the low wage countries.

Theory and the doubters' arguments fail equally with the low wage countries. During the first decades after the War many of these countries tried to industrialise by investing in and protecting new industries, particularly to produce capital goods, motor vehicles and the like, industries which rarely became competitive although the economic growth rates may have been high. The exceptions are economies like Taiwan, South Korea

and China¹ (in chronological order), of which the firms have become competitive in these and other technically advanced goods. But they became so by acquiring by one means or another technical knowledge from the firms of the high wage countries, something outside the scope of orthodox theory; nobody claims these economies' acquisition of technical knowledge conformed to their comparative advantages. Countries like Malaysia and Thailand, which now produce some technologically advanced goods, may seem to be exceptions, too, but are merely the locations for the firms of high wage countries that produce those goods.

These points are not meant to imply that international trade is impossible to explain. They can be adequately dealt with by starting from the right premises and this work is an attempt to show that much of international trade can be plausibly and rigorously explained by straightforward reasoning from premises that few would dispute.

An explanation of trade must first distinguish between trade that arises from differences between countries and that which does not. Comparative advantage is meant to explain the former and concludes that, except under special circumstances, countries export goods different to those they import. It is the lesser part of trade that is explained like this; the greater part consists of exchanges of similar goods, often between economies that are also similar, and must be explained differently. Chapter 1 is an attempt to show how trade does arise from differences, but differences of wages. Chapter 2 is an attempt to explain trade as the consequence of the R&D of firms and the resulting technical progress. In this kind of trade countries can exchange the same goods. It also explains the varying degrees of success of the economies of Taiwan, South Korea, China and India. It leads, moreover, to the question of the effects of exchange rate changes on prices and to discussion of the balance of trade in Chapter 3. This may seem odd since the balance of trade and the determinants of trade are usually discussed separately with different sets of assumptions. But, if countries exchange similar goods in competitive markets, the usual arguments about the effects of exchange rate changes cannot be used. Instead, the reasoning has to rely on the accounting relations between the trade balances, expenditure and production of the various countries. Finally, Chapter 4 describes the early years of Pakistan's economy to show the consequences of following what were supposed to be its comparative advantages. What happened in those early years has determined nearly all that has followed and is illustrative of what has happened in many low wage countries.

¹The country, China, is regarded here as comprising three economic systems, which will be referred to here as China, Hong-Kong and Taiwan.

Of the various differences between countries the one of interest here is wage differences. They must be taken as a cause of trade, even by those who believe them to be determined by comparative advantages; they have been too great and too long-lasting not to have influenced investment and production. Chapter 1 describes how they can give rise to trade when countries are similar in all other respects and uses a simple model to demonstrate the consistency of the reasoning and bring out points easily missed in a purely verbal exposition. It also discusses explanations of trade by other kinds of differences between countries, the Heckscher-Ohlin theory in particular, and shows that, when not trite, their reasoning is fallacious.

There are three other kinds of differences that are used to explain trade. One is the natural conditions of each country, such as mineral deposits, climate, soil, forests, sea coasts, lakes, rivers and so on. They account for a part of international trade, though not a great part. Ricardo's theory can be understood as an attempt to explain trade by such differences. But is a theory needed to explain why Kuwait exports oil or Ghana cocoa? This part of trade is taken for granted and need not be discussed further.

The second kind of difference between countries is that of the proportions of the factors with which countries are endowed and its use for explaining trade is the Heckscher-Ohlin theory, which has been standard for a while. Chapter 1 shows the theory has flaws both in its assumptions and in its reasoning. Its most important flaw is its neglect of produced factors. Its factors include natural conditions of the sort just mentioned, but it must also include factors that are produced, not natural. But here it fails; produced factors of production evidently include capital goods, which are heterogeneous, and each must be made in the country where the factor prices allow it to be made most cheaply. Hence the prices of capital goods are determined by the prices of factors everywhere. There is a dual to this; each capital good installed in a country is installed there because it is expected to yield more profit than one that has not been installed there, and that profit also depends on the prices of goods and factors there and elsewhere. A country's endowment of a produced factor of production, therefore, depends on factor prices, and capital goods cannot be aggregated into one factor, capital, as is the practice with the Heckscher-Ohlin theory.

A third type of difference is that countries do not all have access to the same techniques of production. A method of producing a particular good in one country may not be usable or give the same results in another, even though the equipment used, the amounts and kinds of labour and all other conditions are the same. This type of difference must be excluded from the Heckscher-Ohlin theory, or else production techniques can be postulated to suit virtually any combination of factor endowments and trade and the theory

becomes vacuous. In practice, though, much the same manufacturing processes can be used in different parts of the world, even if they have to be modified to allow for climatic conditions, such as heat or snow, and may be impractical in countries that lack water. But these are minor qualifications and, unless there is a specific reason for taking them into account, it is assumed here that a production technique used in one country can be used in another with the same results, assuming that the country has the technical knowledge and trained workers needed.

To see what the consequences of wage differences alone can be, the first step is to take nominal wages as given and assume away other causes of trade; to assume, therefore, that all countries have equal access to the same production techniques and that returns to scale are constant. With free trade and the conventional assumption that transport costs are negligible, all countries obtain the same prices and, therefore, income from producing the same goods. Where the wage is lower profit margins are higher. So, a country with a lower wage produces the good that yields the highest profit rate there, which means that other goods cost more to produce there because the high profit offsets the low wage. For example, that country may produce cloth, but import looms from a higher wage country, where looms cost less to make because the rate of profit is lower.

Low wage countries in which there is not enough capital equipment to employ all the labour available can increase production by adding to their capital stocks by importing or producing locally, whilst wages can rise in those with no unemployed labour, which does not mean they will. It can be expected that, when lower wage countries increase their production of a good also produced by countries with higher wages, it is the latter that must reduce their output of that good if the total output has to be reduced, for the former countries can always lower their prices until the profit margins of the latter are zero. But the higher wage countries can avoid unemployment by adjusting their production to the demand for capital and consumer goods of the lower wage countries. Free trade benefits the low wage countries that have unemployed labour by increasing both wage and profit income and, beyond the possible costs of adjustment, causes no loss to the other countries.

But, if a group of high wage countries avoid or delay adjustment by uniting to raise barriers against imports from the other countries, though not from one another, and thus lower the prices of the lower wage countries' exports, they transfer some of the gains from trade to themselves in the form of lower prices and tariff revenue. There is a limit to that, though. In the lower wage countries lower prices yield smaller profit margins and, taken far enough, makes some goods that were imported from the high wage countries cheaper to make domestically. Thus, if a lower wage country imports a

capital good from a high wage country when there are no trade barriers and then trade barriers reduce the rate of profit by enough, it becomes cheaper to make that good domestically. Taken far enough such barriers lower the rates of profit of the low wage countries to the point at which their highest profit rates come from making all their capital goods themselves with nothing to gain from trade.

This reasoning is set out in Chapter 1, which goes on to show that the Heckscher-Ohlin theory cannot possibly hold, even leaving aside the point already made about its incompatibility with produced capital goods. It shows that a factor cannot in general be disaggregated into several factors, or the reverse, and that there must, therefore, be a single, definitive list of factors for the whole world if the theory can be claimed to explain trade. Proponents of the theory do not have such a list and have shown no sign of wanting one; they rather keep their factors arbitrary. At the simplest there are two factors, labour and either land or capital, but the number is often not specified. Similar reasoning holds for the products. Staying in the abstract like this allows much of what passes as the theory of international trade to be theorems on the properties of positive matrices and their principal minors. Nevertheless, there are points at which comparisons with reality are unavoidable, and there the theory fails. It concludes that countries have the same factor prices when they produce more goods in common than there are factors, but that each exports goods different to those it imports if their factor prices are not the same. It also failed the classic empirical test of the theory by Leontief, which showed that the exports of the US, the country with the greatest relative endowment of capital, were relatively labour intensive. Later versions of the test have yielded diverse results and are no evidence for the theory.

If trade, free or not, is determined by the given wage rates, it cannot be presumed to allocate resources efficiently. In the Ricardian and Heckscher-Ohlin theories productivity and endowments are given, along with preferences, wages and prices are the consequences and resources are allocated efficiently. On this depends these theories' appeal. Changes of productivity and endowments affect the future but not the present. It follows that, instead of protecting the agriculture of the temperate zones when rendered uncompetitive by high wages, although it could produce more, it is supposedly more efficient to abandon the protection and leave the land fallow while the Amazon burns to provide for the consumption of these zones.

In practice, the low wage countries have not reacted in the way described in Chapter 1 to the trade barriers the high wage countries did impose. Practically no low wage country has started to produce its own capital goods because of these barriers. Countries like Pakistan and Bangladesh, of which textile and clothing exports have been continually

subject to tariffs and quotas, import practically all their capital equipment. India produces much of its own capital equipment, but set out to do so from the start, on becoming independent, and Taiwan, South Korea and China, which can now produce goods as technically advanced as those of the high wage countries, grew mainly by diversifying to other industries rather than industrialising by producing the looms etc. for their textile industries.

What needs explaining is why so few low wage countries have tried to deal with the trade barriers against their exports of the high wage countries by producing their own capital goods. Why, for instance, when the high wage countries made their various "Arrangements" and "Agreements" to restrict imports of textiles from low wage countries, did the latter not produce their own spinning machines and looms? Since it takes capital as a given endowment of a country, the Heckscher-Ohlin theory cannot even pose the question, and it follows from what has been said already that the infant industry argument does not provide an answer either. When the question is put like this, a non-economist would answer that capital goods in the form of machines are more difficult and complicated to make than most textiles and garments, even if these, too, need expertise. It is a reasonable answer, but insufficient, for, if all countries have equal access to the same techniques of production, a country need only make the same investment in training workers and engineers as the high wage countries to be able to produce the same machinery, again with lower wages and higher profit rates.

To this the non-economist would add that, with the exceptions already mentioned, the low wage countries do not have the technical knowledge of the high wage countries. This, too, is a reasonable answer and is formalised in Chapter 2, which discusses technical progress, by dividing goods into two categories. One is the category of "complex goods", which are constantly being improved through the research and development (R&D) of the firms that make them. Goods of this category, like motor cars, skis and bioreactors, are differentiated, which is to say that there are at any time different versions, differing in their characteristics and prices to meet the demand of buyers with different preferences and budgets. As new, improved versions come onto the market, the prices of older versions fall and firms withdraw these versions to avoid losses or to release production capacity and labour for making their newer versions. Technical knowledge, therefore, can be divided into that generated by the R&D of firms, which is proprietary knowledge protected by patents and secrecy, and generally available knowledge, including the knowledge made available as patents expire. "Simple goods", the other category, are not changed by R&D. Examples are most cotton textiles, many garments, simple implements and tropical and sub-tropical crops. The division is not permanent; even ordinary crops may have to compete with newly bred or genetically modified versions.

Low wage country firms can use generally available knowledge to make versions of complex goods that are out of date by high wage country standards and may be able to compete because their low wages allow low prices. But they cannot expect to catch up with the technical knowledge of the high wage country firms, which have a lead and continue to invest in R&D. A low wage country firm has no interest, therefore, in incurring the cost of R&D to generate its own proprietary knowledge, unless, like in India, it is protected against foreign competition, with the consequence that its products stay technically behind those of the high wage country firms. What Taiwan, Korea and China did was acquire the advanced knowledge of high wage country firms by means such as subcontracting and joint ventures, and their higher education provided enough adequately trained scientists and engineers for their firms to be suitable partners for the high wage country firms and for some of them eventually to do their own R&D.

A comparison of a developed economy of fifty years ago with the same economy now shows the limitations of the free trade and infant industry arguments. Few of the complex goods of the old economy would compete with up to date versions except at prices that would not cover their production costs. And some goods produced now did not exist then. Yet the economy of fifty years ago was a viable economy. Viability, therefore, does not imply competitiveness; an economy can be viable with protection but not competitive because its firms do not have sufficiently up to date technical knowledge and are unlikely to catch up by themselves with the firms of the high wage countries.

There can be any number of versions of a complex good and, because producers cannot copy or imitate the versions of other producers' versions without permission, competition ensures that several are on the market at any time. Producers in different countries of the same good compete in the same markets with their different versions and a country can import the same good as it exports. Firms can price the same versions of goods differently in different countries and arbitrage is prevented from eliminating the price differences by the laws on intellectual property rights, which allow agreements of producers, importers, wholesalers and retailers that exclude third parties. World markets only exist for some undifferentiated goods and the common depiction of competition as taking place in one is misleading. It is, therefore, to be expected that prices converge slowly and that changes of exchange rates increase or reverse the convergence from time to time, and that is what econometric comparisons of the price indices of different countries seem to show.

Part of the price differences between countries comes from local costs. Standard depictions of international trade are also misleading in assuming

away the distinction between the final price and the factory gate price of a locally made good or the price paid to the exporting country of an import. In between are the costs of bringing the good to final sale, costs that are often the greater part of the final price. With an import they include, apart from the shipping, several local costs, such as dock charges, storage, transport, insurance, sundry fees and taxes, all of which can be taken as given, and the profit margins of the importers, wholesalers and retailers who cover these costs. These are costs of untradables that vary with the good and the circumstances. Some of them are not incurred for goods ordered by the final users directly from the producers, usually capital goods, whereas retailers of consumer goods include in their final prices the cost of sales staff, rent of premises and packaging.

It cannot, therefore, be said *a priori* how a change of exchange rate affects imports or exports. Demand depends on the final price, which may or may not change when the exchange rate changes, and the effect on the trade balance depends on the price received by the exporting country, and in between are the local costs. At one extreme the price received by the exporting country can stay constant in terms of that country's currency, in which case the profits of the importers etc. make all the adjustment, and at the other it can be unchanged in terms of the importing country's currency. Producers of differentiated goods usually have agreements, varying according to the goods involved, with their distributors in their export markets and these, along with competition in the importing country, determine where the outcome lies in this range.

Even the standard assumptions of homogeneous goods, no local costs and competitive world markets, do not give a relation between a change of exchange rate, on the one hand, and imports and exports, on the other. Competition means one price; if, because of devaluation, producers in one country lower their price of a good, their competitors elsewhere do the same. Only if the devaluing country has unemployed labour and spare capacity can its output of that good increase without reducing its output of other goods, and that increase depends on how much world demand is raised by the lower price and how much competitors reduce their output and employment. If there is such an increase, income from production may also increase and, then, so may expenditure on imports and on goods that would otherwise have been exported.

Income from production, which is also the value of production, is to be distinguished from expenditure, and the trade balance is the amount by which the former exceeds the latter. This and the equality of the sum of all countries' trade balances at any moment with zero are identities and inescapable. In all countries the residents have *ex ante* intentions of

expenditure and firms have production plans, but there is no reason that the trade balances they imply should be mutually consistent. As an extreme example, each country's intended expenditure may be less than the production plans of the firms, implying trade surpluses for all.

In the outcome these inconsistencies are resolved and Chapter 3, on the balance of payments, discusses how. It is convenient to assume that each country's total expenditure is determined by the monetary authorities and that production adjusts to resolve any inconsistencies, though the reasoning can be adapted to other assumptions. Some countries do better than others in that, *ex post*, production by their firms is closer to the *ex ante* plans. In principle, all countries can agree on a mutually consistent set of trade balances and coordinate their expenditure accordingly, though there never has been an attempt at such an agreement. But it would not, by itself, ensure that all countries produce to the limits of their production capabilities. This is where the exchange rate is useful; by devaluing a country lowers its wage rate relative to prices elsewhere and perhaps brings unused capacity into production. If all countries agree on the trade balances, the devaluing country can expect that the others accommodate its increased production as long as its expenditure conforms to its agreed trade balance. Only to this extent is there assurance of the efficacy of devaluation.

Can there be an automatic coordinating mechanism to bring about mutually consistent trade balances? If the market were to do it on its own, it would have to determine both all exchange rates and expenditure in each country. This is not the standard argument for letting exchange rates float, which merely presupposes a determinate relation between a country's exports and imports, on the one hand, and the exchange rate or change of exchange rate, on the other, and leaves out expenditure.

Such a relation, were one to exist, would have to be in some way inherent to the economy and, therefore, independent of nominal prices. From it follows the notion of a correct exchange rate in free trade equilibrium, independent of nominal prices and inherent to the economy. Definitions and methods of estimating it abound and more are constantly being devised. Among them are various shadow and other exchange rates derived from various optimisation and general equilibrium models and a variety of Real Equilibrium Exchange Rates related to some hypothetical equilibrium. Their properties vary; some of the former supposedly ensure the correct balance of trade and have powers of resource allocation that verge on the mystical, whereas some of the latter are not even the exchange rates at which the markets for currencies clear. Their variety alone gives grounds to doubt the validity of any; if the notion of a correct exchange rate were valid there would be only one such rate. Part of Chapter 3 is devoted to showing the fallacies involved.

A coordinating mechanism of sorts was provided by the gold standard as it became nigh universal during the half century before the First World War. In theory it made the *ex ante* macroeconomic balances of the different countries roughly consistent by obliging the countries with persistent trade deficits to reduce expenditure as their gold stocks diminished and letting those with surpluses increase theirs. Practice was not as crude but the complexities it had to cope with grew with time. There were long periods during which the gold supply failed to keep up with economic growth, periods of falling prices, countries could and sometimes did hoard gold instead of increasing expenditure and changing the price of gold was ruled out. Recourse to financing devices circumventing the normal gold standard mechanism became more frequent and were a privilege available to countries like Britain, but not to countries of Latin America, which often had balance of payments crises. Perhaps the system would not have lasted much longer had the war not brought it to an end.

One reason it lasted as long as it did was the confidence in it; countries running out of gold and other reserves, at least industrial countries, could be expected to raise interest rates and restrict expenditure as much as needed if the currency depreciated to the gold point. Speculative capital was attracted to such a country because of this confidence and added to its reserves. Once that confidence was gone it proved impossible to restore and attempts to return to the gold standard after the war were bound to fail, for a change of exchange rate was now a possibility and caused speculative capital to flow from deficit to surplus countries. This, with what can now be seen as other economic mistakes and the political conflicts of the time, and turmoil was assured until the Second World War ended.

The Bretton Woods system adopted at the end of the war was unique in that it was a system designed and put in place by agreement. It was akin to the classical gold standard in that exchange rates were fixed and that the main means of international payment, now the US dollar, was on the gold standard, though confidence that the US would let its gold reserves constrain its money supply when the time came did not last long. Like the gold standard, the gold exchange standard had the drawback of putting no obligation on countries with persistent surpluses to increase expenditure. It also explicitly allowed countries with persistent deficits to devalue as a last resort, so that speculative capital flowed from deficit to surplus countries and could oblige a country to devalue before its reserves ran out. Eventually it collapsed as devaluations became more frequent, though another cause was the growing belief in the desirability of floating exchange rates. Nevertheless, the period before its collapse was, despite the volatility of foreign exchange markets, one of economic stability, growth and evenness of income distribution among the high wage countries that has never been equalled.

At present the orthodoxy is that currencies should float, though the doctrine behind it is obscure. On the one hand the exchange rate is supposed to indicate something about the economy, though it is not clear what, and, on the other, several countries, especially in Europe, where orthodoxy prevails, have persisted in trying to fix exchange rates among themselves, to the point of adopting a common currency. Again, the US dollar is the main international currency, and, as long as countries and markets continue to accept it as such, the US can run current and capital account deficits freely. It had run deficits on the capital account during the time of the Bretton Woods system, turning the “dollar shortage” into a “dollar glut”, and thereby provided much of the liquidity for the speculative capital that ended the fixed exchange rates. An attempt to provide an alternative banking reserve, the Special Drawing Rights, has depended on approval of the US and the countries of Europe and has remained insignificant. But the system seems stable because central banks are not concerned with the real cost in terms of output transferred to other countries and firms and individuals can buy assets in the US and elsewhere practically without constraint.

There are several reasons for discussing the balance of trade and the exchange rate at such length. One is to show that taking capital goods as durable and heterogeneous allows these matters to be discussed using the same assumptions about production as the discussion of the pattern of trade and production, in contrast to the usual practice of using two separate sets of assumptions. Another is to show that it follows from the accounting identities that trade balances are the outcomes from the mutually inconsistent expenditure and production plans of the residents of the various countries. Thirdly, exchange rate changes cause relative changes of nominal wages and have, with few exceptions, systematically lowered the nominal wages of low wage countries in terms of the currencies of the high wage countries. Fourthly, when exchange rates are left to be determined by the market, the mechanism by which the hard currencies of high wage countries are determined is not the same as that of the currencies of low wage countries.

Chapter 4 is an attempt to show how the arguments given here explain Pakistan’s economic development and illustrate much of what has happened in other countries. It is argued in Chapter 1 that, if the high wage countries impose tariffs on their imports from the low wage countries, they reduce the prices the low wage countries receive. This is what they did to Pakistan’s exports when it began to industrialise in the 1950s and 1960s; they imposed tariffs and other restrictions on its exports of cotton and jute manufactures, but not on the raw cotton and jute that it also exported and which the importing countries did not produce. Pakistan’s value added, hence income, from manufacturing was lowered and so, therefore, was its ratio to the cost of the investment for the manufacture. Cotton and jute manufacturing had

low capital to output ratios at the prices of the high wage countries, but high ones in Pakistan. These industries seemed to be labour intensive, but were in practice capital intensive and gave low economic yields.

By the argument of Chapter 1, Pakistani firms would have started producing capital goods if they could. By the argument of Chapter 2 they lacked the up to date technical knowledge to produce competitive ones. Pakistani firms could, perhaps, have exported profitably using uncompetitive capital goods made locally with out of date designs. That is to say Pakistan could have done the same as India and protect its capital goods industries. But the government and its foreign advisers believed that capital goods did not conform to Pakistan's comparative advantages, which were limited to cotton and jute textiles and similar simple goods, none of which required much education. Education in general, however desirable in itself, was of low priority in the pursuit of comparative advantage, and the prerequisite for a capital goods industry, an appropriate system of higher education in the natural sciences and engineering, of even lower priority.

At the time Pakistan's economic policies were much praised; the economy was even described as about to take off and foreign aid was abundant. For the government and officials pledges of foreign aid were a sign of approval; the more their policies were approved, the more foreign aid they could hope for. Debt service increased relative to exports and probably cost as much foreign exchange as was gained from the investments in cotton and jute manufacture. Since then, with a brief interruption in 1972-76, successive governments have continued conforming to orthodox doctrine. Import duties have been reduced, losing thereby a large part of government revenue, which governments have been trying to replace, with modest results, by taxes on income and sales. Government revenue is too little to improve education, health and other social services, which rank among the worst in the world. The currency has been allowed to depreciate from Rs. 9.61 to the dollar in 1972 to about Rs. 160 now, without any noticeable improvement of exports; the goods produced now are much the same as at the end of the 1960s and the nominal wage has increased less in terms of the dollar between 1960 and now than US prices. Transfers of capital abroad are legal and easy, leaving no restraint on capital flight, which the depreciation stimulates. Yet the external debt is so great that it has to be rescheduled when there are untoward events.

It would be wrong to dismiss this chapter as mere history. An economy can only be understood as the outcome of its history and neither history nor the passage of time has had much influence on orthodox economic doctrine. One set of observations from history is that no economy has industrialised without protection, no industrial economy refrained from protection against competing exports from low wage countries and none nowadays dispenses

with state support or the occasional use of trade barriers. Such explanations as orthodox theory can provide, especially altering the terms of trade and returns to scale, are implausible and do not explain the high universality of the observations. Proponents of orthodox theory often prefer to assert that the industrial countries would have fared better had they avoided protection, but, by the same token, they cannot explain why there was protection. Another observation is that since the mid-nineteenth century, at least, rising wages have been a stimulus to increasing the output of workers and improving consumer goods in the high wage countries. Keeping wages down relative to the cost of capital equipment, which is the purpose of devaluation, removes that. Orthodox theory cannot both explain the one and justify the other without splitting into two or more theories for two or more sets of countries.

Numerous modifications of orthodox theory have been made to adapt it to one aberrant phenomenon or another. But the theory is limited by the notion of capital on which it depends and its consequent reliance on production functions. These allow some choice of parameters and the numbers of goods and factors, and adepts construct various models to fit different phenomena. They demonstrate that the theory is malleable, but also that it is not explanatory; *ad hoc* adaptations are not explanations unless corroborated by other evidence. Were economists to construct a model that included all the aberrant phenomena it might be argued that there was reason to seek corroborating evidence. But there has been no such comprehensive model; the assumptions made for matching one phenomenon are inconsistent with the assumptions for another.

What follows can be put under three heads; theory, history and criticism. Theory is the reasoning regarding international trade and the balance of payments set out in Chapter 1, Section 2, Chapter 2, Sections 1 and 2, and Chapter 3, Section 1. Second is exposition of some economic history in Chapter 1, Section 1, Chapter 2, Section 3, Chapter 3, Sections 2 and 3, and all of Chapter 4. Third is criticism of the relevant orthodox theory and various models in Chapter 1, Sections 3, 4 and 5, Chapter 2, Section 1, and Chapter 3, Sections 2 and 3.

The reasoning under the first head can be read by itself. It is all simple or common sense; there is no grand theory. Some of the conclusions reached here differ from established ideas and, though cogent reasoning from accepted premises may suffice in terms of pure theory, its validity must be judged empirically and it is more persuasive if the reasons for the differences are explained. Empirical support comes mainly under the head of history; it accords with what has been observed and explains crucial elements of the economic successes and failures of various low wage countries, though some of what may seem to be the excessive space devoted to history is this work's argument for fixed exchange rates.

One reason for the differences is, as pointed out earlier, that the notion of capital on which orthodox theory depends cannot accommodate trade in capital goods. It is part of an older and broader objection to quantities of capital and production functions set out in the appendix. A second is that the argumentation of orthodox theory has many fallacies, some of which could not have escaped their authors. Even if the fallacies have been exposed before by others, they must go on being exposed as long as they have currency.

Acknowledgements

Directly or indirectly almost all the argument of this work has followed from the rejection of the notion of capital in production as a quantifiable mass in favour of considering it as what it really is, a heterogeneous collection of goods, many durable, that were, themselves, produced by using such goods. My first thanks go to Amit Bhaduri, who was the one who made me aware of the issue. It was over a long walk on the backs of Cambridge at the time the dispute over the notion of capital was at its height, that Amit, himself a prominent participant in the dispute, explained to me what was at issue.

I also owe a debt to Prof. David Field, who programmed Titan, the Cambridge computer, to do the calculations referred to in Chapter 4 using the formula for evaluating investments in Chapter 3, and who provided comments that have helped make the text more readable. Another debt is to Prof. Jon Wisman, who, in addition to comments that have improved the text, provided or made me aware of publications that I would otherwise have missed. The Lahore School of Economics gave me the opportunity to teach, and, therefore, learn, there over several years and to put in print some of the ideas of this work. It is also thanks to the Lahore School's yeoman work in putting historical economic data for Pakistan online that I could put numbers to some of the last chapter.

Finally, I must acknowledge my debt to the World Bank, where I worked for almost twenty years. It was there that, apart from acquiring a wider acquaintance with the economies of low wage countries, I learnt the rudiments of national and balance of payments accounting, banking and prudential regulation, direct and indirect taxation, international trade regulations and various other subjects that should be the stock in trade of the practical economist but are rarely treated as concrete matters in economic textbooks. Despite its doctrinal biases and the political interference, it was once an impressive institution.

None of those mentioned should be held to share any of the views expressed in this work or to be responsible for any of its errors and shortcomings.

CHAPTER 1

COMPARATIVE ADVANTAGES AND CAPITAL GOODS

1. INTRODUCTION

When Adam Smith in “The Wealth of Nations”² in 1776 and Robert Torrens in his “Essay on the External Corn Trade” of 1815,³ among others, asserted that there was an economic advantage to buying an import if it were cheaper than the equivalent domestically made good they did not ask the question as to why some goods should be produced more cheaply in some countries than in others. Nor did they ask what the secondary effects of importing and exporting would be on, for instance, the overall balance of trade or the composition of country’s production. Their arguments were more in the nature of common sense than a comprehensive theory and were intended to counteract the mercantilist belief in the virtue of accumulating gold through trade surpluses. Smith, moreover, insisted on the futility of trying to prohibit imports or exports.

Ricardo’s theory of comparative advantage, enunciated in his “Principles” of 1817, provided an answer to these questions that seemed for a long time so comprehensive that it remained the orthodoxy until 1933, when Ohlin published his elaboration of the factor endowments theory⁴, originally thought of by Heckscher in 1919, according to which trade was determined by the relative abundance or scarcity of factors. The comprehensiveness of Ricardo’s theory was that the same mechanism both brought the trade into balance and led countries to an optimal pattern of production, that is to their comparative advantages, and the mechanism was automatic. Taking Ricardo’s own example, if England required the labour of 100 men for one year to produce a certain quantity of cloth and the labour of 120 men for one year to produce a certain quantity of wine, whilst the labour that Portugal required for the same quantities of cloth and wine was 80 men and 90 men for a year, England would never export wine at the same

² Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations: Vol. 1-2*.

³ Torrens, *An Essay on the External Corn Trade*, . 1st and 2nd editions. The third edition conforms to Ricardo.

⁴ Ohlin, *Interregional and International Trade*.

time as Portugal exported cloth. Cloth would be cheaper to make in England than in Portugal or wine would be cheaper to produce in Portugal than in England. Either one of the countries exported both goods, or England exported cloth and Portugal wine. In the former case the trade surplus of the exporting country and the deficit of the other caused gold to be transferred to the latter from the former. Prices, then, rose in the former and fell in the latter and, since prices in the former must at the start have been lower than in the latter, the latter's price of one good eventually fell to the level in the surplus country, at which point the deficit country began to produce that good. The other might be producing both goods, but, if the trade imbalance and gold transfers continued, the price in the deficit country of the good it was producing fell below the cost of production of that good in the other country. In the end one of the countries produced one good, which did not prevent the other country from producing both goods, though it was more likely that each country produced one good, England cloth and Portugal wine, and that the relative prices lay somewhere in the range between the relative costs of England and of Portugal.

Ricardo's theory was comprehensive in a way that the writings of Smith and others had not been. Where others had taken for granted that goods might be imported because they were cheaper to produce elsewhere, Ricardo took account of the repercussions. Beginning with specific characteristics of countries, namely the amounts of labour needed to produce the various goods, the pattern of production and trade, i.e. which goods each country made and which it imported and which it exported, was determined by how these characteristics compared. The whole economy of each country adapted to trade and that set the prices. John Stuart Mill showed how the price indeterminacy left could be removed by adding demand as a function of price. Mill, Mangoldt and Edgeworth also extended Ricardo's reasoning to several goods and others to several countries.

It was this comprehensiveness, along with the desirability of the outcome, that made it so persuasive for over a century. One reason the outcome was desirable was that it was optimal: both countries gained from trade, for each obtained more goods with trade than without, and the gain was all that was possible, for if one country were to gain yet more it would have had to be through the other gaining less. This followed from Ricardo's theory of value, which asserted that the prices of goods produced in a country were proportional to the amounts of labour needed to make them. Another reason for the outcome's desirability was that it came about automatically through the workings of unhindered commerce and seemed to leave no grounds for government interference, which accorded well with the spirit of an age when, in Britain especially, such interference was regarded as likely to be ill-judged or to favour special interests. Britain's pre-eminence as the

most advanced industrial economy in the world added to the prestige that the “Wealth of Nations” had in any case acquired as part of progressive, liberal thought in Europe, and the harmony of spirit, if not of detail, of Ricardo’s theory of trade and Adam Smith’s teachings made free trade the orthodox doctrine of economists.

But, though comprehensive, the theory was static and it was this that aroused the first objections. It assigned to a country the economic activities to which it was supposedly best suited and implied there were no better alternatives. Portugal would continue to produce wine and England cloth. But, despite Ricardo’s care in assuming in his example that Portugal could produce both more wine and more cloth per head than England, it was evident that progress came through industry, not wine. And, if that was the case, it seemed paradoxical that, if Portugal had been less efficient at producing wine, needing, say, the labour of 100 men, instead of 90, for a year to produce the amount of wine that England produced with 120 men, it would have been producing cloth and England wine. The combined income of the two countries would have been less because of the lower output per head of wine, but, outside of Ricardo’s static framework, this might have been preferable for Portugal.

Thus began a conflict that has continued with varying degrees of intensity between the proponents of free trade and those who objected that the theory was static. The objectors who had to argue against a theory that provided so plausible a framework for thought on international trade and that concluded that trade was best left to itself had to argue from outside the framework. When Alexander Hamilton in America expounded his “infant industries” argument, that countries needed to protect their new industries against the competition of established foreign producers through the early stages if they were to have any industry at all, it was before Ricardo had formulated his theory. He was aware that the ability to compete with Britain’s industries required at least experience and, therefore, time. In this he was in keeping with his time, for governments often protected local industries, sometimes successfully enough for them to flourish later as exporters. Daniel Defoe, writing in 1730, described how England’s wool manufacture had acquired the ability to produce cloth that could compete with the products of Flanders because of protection under Henry VII and Elizabeth I. Similarly, England’s cotton industry was protected against India’s in the eighteenth century until machine design had advanced enough for English cottons not to need protection.⁵

Hamilton’s view prevailed in America even after comparative advantages had become the orthodox doctrine of political economists.

⁵ Chang, *Kicking Away the Ladder: Development Strategy in Historical Perspective*, 19–21.

Friedrich List, who had lived in America from 1825 to 1830, with a brief interruption, elaborated the arguments American proponents of protection, like Henry Clay and Daniel Raymond, had derived from Hamilton and advocated policies similar to theirs to the German states until his death in 1846. Like the Americans who influenced him, his objection was to the static nature of Smith's advocacy of free trade, which was then the most influential writing against mercantilism, though not as coherent as that of Ricardo, whom List barely mentioned. He substantiated his argument with examples of industries, including those of England, that flourished after having been protected at the start and failed otherwise. Over the nineteenth century both the US and some German states industrialised while protecting their industries against British competition. Portugal was less successful; when its government tried to protect its cloth producers it received lectures on comparative advantages and the benefits of free trade from the British Foreign Secretary, Palmerston, while the British navy ensured compliance. Palmerston gave some German states the same lecture, though they worried less about his navy. But List did believe the British secret service financed letters and pamphlets in the German states in favour of free trade.⁶

The comprehensiveness of Ricardo's theory also posed a dilemma for later political economists whose theories of prices were not that of Ricardo; unless they were to be agnostic, they had to choose between inventing new trade theories consistent with their price theories and accepting Ricardo's theory despite the inconsistency. Ricardo, himself, was aware that prices were not strictly proportional to amounts of labour needed for making units of goods because of the element of profit, but he presumably considered these variations to be negligible. Such was the appeal of comparative advantages that nearly all others who wrote on trade preferred inconsistency to agnosticism, though they tried to find ways of showing that their prices corresponded inversely to possible output as required by Ricardo's theory. They varied in how they formulated it, but most assumed that the profit element was a cost, a kind of disutility that could be added to the cost of labour, in Senior's formulation a sum of 'labour and abstinence' and what Viner called the 'real' cost.⁷ Ricardo's presumed assumption that the element of profit in cost made little difference, either because it was small or because it varied little between goods, was used by Taussig and Viner. Despite the awareness that there was no coherent formulation of how prices and output could be related to satisfy its requirements, Ricardo's theory remained the orthodoxy. Even in the 20th century Pigou, whose theory of prices was that of Marshall, wrote a defence of free trade that relied on Ricardo.⁸

⁶ List, *Das Nationale System Der Politischen Oekonomie*.

⁷ Viner, *Studies in the Theory of International Trade*, chap. VIII. Section 9.

⁸ Pigou, *Protective and Preferential Import Duties*.

Heckscher and Ohlin resolved the dilemma more than a century after Ricardo first proposed his theory by inventing a new theory. In 1919 Heckscher surmised that the pattern of production and trade might be determined by the relative abundance or scarcity factors of production in each country. It was the first real break with Ricardo's theory and was given a fuller exposition by Ohlin in his work, "Interregional and International Trade", in 1933. Thus, if Sweden had more forest compared to land, labour, capital and any other productive factors than other countries, it would export goods that required more forestry products, like lumber. Similarly, if North America had wheat lands in greater abundance in relation to forests, capital, labour and so on than other countries, it would export wheat. Ohlin gave several other examples with factors as diverse as mines and sixteenth century craftsmen.

In place of the single resource, labour, of Ricardo there were several factors, in which rough form the theory seemed to fit the prevailing views on production and prices deriving from Marshall and Walras. The addition of several factors, that is the notion of production functions, which give the amounts of goods produced for any combination of the several factors, seemed to present no problem. At first sight the theory seemed to enable economists to explain international trade with a degree of concreteness that Ricardo's theory lacked; factors were thought to be observable and production functions to represent known technical relations.

Like Ricardo's theory, the factor endowments theory is comprehensive and concludes that, under free trade, the workings of competition reach a stable, optimal outcome without need for government interference. Its justification of free trade is the same as that of Ricardo's, namely that each country's production and trade is then optimal in the sense that the quantities of goods available to the country are at least as great as they can be otherwise, with the proviso that no country can use monopoly power over some tradable goods to set their prices by itself. The conditions, especially regarding the production functions, that allow the conclusion are normal for economic theory. Where the factor endowments theory seems to be an improvement over Ricardo's is that it avoids unexplained differences of productivity; all countries can have the same production functions, but they have different endowments of factors. It also seemed at first that the more abundant a factor was relative to other factors in a country the lower was its price there and countries with little capital per head had low wages and high profit rates.

Both theories are theories of comparative advantage in that they assert that the whole of each economy adapts to foreign trade according to the characteristics of the country, such as labour costs or factor endowments, and does so by adjusting production and the prices of goods and of factors, including wages and profit, and also assert that, with the proviso above, the

outcome is optimal. Where they differ is in how trade affects factor prices. In Ricardo's theory the wage is the only factor price and it is fully determined. Ohlin argued from the factor endowments theory that trade would normally reduce the extent to which the ratio of any two factor prices differed from country to country. For example, the price of a factor that was relatively abundant in one country would be relatively low there, that country would produce more cheaply those goods of which the production used the factor relatively intensively, it would export those goods and demand in the country for that factor would increase. It seems that Heckscher believed that trade must eliminate differences between countries of factor prices if they had access to the same production techniques. Stated in terms of production functions, there is a set of equations relating the price of each good to the quantities of factors used in its production and to the prices of other goods to the extent that they are inputs in the production of that good. If the number of factors exceeds the number of goods, the prices of factors are not fully determined. If the numbers of goods and factors are equal, the prices of the ones determine the prices of the others, though the equations can have several discrete solutions. If the number of goods exceeds the number of factors, there need not be a set of factor prices corresponding to a given set of prices of goods, though it can be assumed that free trade results in prices of goods such that there will be.

As with Ricardo, the factor endowments theory's comprehensiveness led to the objection that it consigned countries to particular activities and was static. Low wage countries were confined to agriculture, handicrafts and labour intensive industries, whilst those with high wages, having more capital per head, would manufacture steel, machinery and electronic devices, the capital intensive goods. Many low wage countries' governments were as little content to limit their economies in this way as the governments of the US and the German states in the nineteenth century had been to observe the comparative advantages of their countries.

In addressing this last point the two theories resemble each other less. *Prima facie* the factor endowments theory precludes the infant industry argument because of the assumption that the same production techniques are available to all countries, whereas in Ricardo's theory trade comes about from the contrary assumption. If, in Ricardo's theory, the amount of labour needed to produce a unit of some good is assumed to decrease, a plausible reason can be experience in producing the good, which is part of the infant industry argument. In such a case the country's income is likely to increase and its comparative advantage may change so that the country can progress to producing goods that yield more income.

In the factor endowments theory a country cannot gain from production not in accord with factor endowments, except if it can change the prices at which goods are traded. If the government of a low wage country brings about the production of capital intensive goods in that country by departing from free trade, it merely lowers the availability of goods there. If the infant industry argument is to be accommodated, the assumption that all countries have the same production functions has to be modified; it becomes, instead, the assumption that the production function of a country is less efficient when the product is new to the country but that it improves with time and experience to become the same function as that of countries that have been producing the good for some time. Even then the comparative advantage of a country with little capital per head is in labour intensive production.

But the factor endowments theory needs the assumption that production functions are the same for all countries. If they are allowed to differ from country to country the theory becomes useless for explaining trade or for prescribing to countries what goods they should produce; production functions can be found to fit whatever happens to be the pattern of production and trade and, if they are allowed to change as a result of experience, there is no general argument for free trade. Then nothing is lost by abandoning factors and production functions altogether to assume, instead, that each economy can be represented as a convex set consisting of the various combinations of goods that it can produce in a period, say a year, with no further assumption about production except that the tangent at the set's boundary represents opportunity costs, i.e. the amount by which the output of some goods have to be reduced to produce one unit more of any particular good. This is the way of representing economies that Haberler proposed to allow discussion of international trade without having to specify how production took place, and it leads to the conclusion that free trade is optimal for a country that cannot alter the prices of traded goods to its advantage.⁹ A country can obtain greater amounts of goods through trade by producing the combination of goods represented by the point on the boundary at which the relative prices given by opportunity costs are those given by trade, which is the point that is reached through competition. Both Ricardo's and the factor endowments theories can be represented in this way.

Abstracting in this way from how goods are produced may serve for advocating free trade, but is static in that it takes a country's ability to produce as given and can provide no indication of how incomes can rise over time. Ohlin objected to Haberler's representation with the remark that, "Such a reasoning explains nothing unless connected with a mutual interdependence price system, and is as different from the doctrine of comparative cost as

⁹ Haberler, *The Theory of International Trade*.

anything can be".¹⁰ Haberler's representation does conform to the criterion for comparative advantage used here, but Ohlin's point was that it does not constitute a theory. In this Ohlin was right and for the same reason it is not termed a theory here either.

Those who wish to keep the infant industries argument can, in principle, still use the Haberler representation by assuming the convex set of production possibilities expands in specified ways. Or they can be more specific and reconcile it with Ricardo's theory by specifying a relation between the labour cost of a good and the amount produced in the past or the time for which it was being produced. They cannot similarly reconcile it with the factor endowments theory.

But the infant industries argument has another aspect, apart from that of becoming competitive on cost, namely that of differences in type or quality of product. Neither Ricardo's nor the factor endowments theory can address it, though Hamilton and List were aware of it, and it evidently accounts for much of international trade. Britain may have been able to produce cloth more cheaply than the US or the German states in the early nineteenth century, but its industries also produced goods that the industries of other countries did not at first produce, such as looms, locomotives and steamships, and when these countries did start to produce them, it was some time before their products were as good as those of Britain.

It is merely an indication of how technical progress changes the kinds of goods available, be they consumption goods, capital goods or intermediate goods. To ignore the quality of goods and new goods is to obscure both what may be the most important part of technical progress and a part of international trade that is needed for any understanding of trade as a whole. Most discussion of technical progress, especially when it is quantitative or uses mathematical formulae, neglects changes of types of output because there is no general method yet of quantifying them. It seems easier to quantify technical progress in the production of goods that do not change, though it depends on having a measure of whatever is used in the production, which, too, can be complicated by including new or better capital goods. New tradable goods raise the questions, where are they made and where do they go? Much has been written about new goods and the "product cycle", but it has not been in terms of comparative advantages. It is doubtful that the factor endowments theory, or even comparative advantages in general, can be extended to accommodate this form of technical progress, for, if the quality of a good varies from country to country, the homogeneity of each good that the theory relies on is lost, and if trade starts in a new good, it

¹⁰ Ohlin, *Interregional and International Trade*.

cannot be supposed that every country can immediately produce that good with the same production function.

2. TRADE WITH CAPITAL GOODS

What follows conforms as much as the inclusion of durable, heterogeneous capital goods allows to the assumptions of the factor endowments theory, in particular that all countries are assumed to have access to the same production techniques, which excludes the infant industries argument. Economic growth in any country, then, is the consequence either of technical progress, which means new techniques of production, or of increases of factor endowments. All new techniques are, by assumption, available to all countries. Factor endowments can change for several reasons, for instance discoveries or exhaustion of mineral deposits, but the discussion is confined to the accumulation of produced factors and, of these, only capital goods. Capital goods are plant and equipment, like machinery and vehicles, as well as infrastructure and items like land brought under cultivation because of new irrigation. Plant and equipment are tradable because they can be made in one country and shipped to another. Infrastructure is not and has to be made where it is used. Capital goods are heterogeneous, meaning that they cannot be transformed into one another, and durable, lasting for several periods. Some may also take several periods to make.

A tradable capital good that is installed in a country was made where it could be made most cheaply, and where that was depended on how wage and profit rates of the various countries compared. Countries' nominal and real wage rates do differ; they have for long been high in some countries and have always been low in others. Regardless of what the reasons for this may be, it must have had effects on production and trade that can be discussed. To start simply, the world is considered as consisting of two countries, which is the usual procedure in international trade theory, and, so, one country, the high wage country, has a higher nominal wage than the other, the low wage country. Other usual simplifications are that all goods are tradable, though capital goods, once installed cannot be moved until they are scrapped, and trade has no costs. Then the difference in the nominal wage is the difference in real wage.

To see how the wage rate can affect trade and production, the high wage country is assumed to be industrialised with full employment and the low wage country to have unemployed workers and no industry at the start. In other words, the high wage country is endowed at the start with a full stock of capital goods and the low wage country with none. Then a good made in the high wage country yields a higher profit margin if made in the low wage

country when the same production technique is used and, if there is a choice of technique, that profit margin is higher still. Hence, when the low wage country starts investing in producing a good that the high wage country produces, it imports the capital equipment for making that good, and the good and technique chosen are those that yield the most profit. This new production displaces the production of the high wage country because the low wage country producers can always lower their price below that of the high wage country producers and still have a higher profit margin. But the high wage country shifts to producing the capital equipment wanted by the low wage country and to adding to the equipment for making it. If the shift is perfectly coordinated, unemployment in the high wage country caused by the shift is merely transitional and the price of the good the low wage country has started to produce rises, instead of falling, because the capital equipment for making it in the high wage country is scrapped earlier and has, therefore, to be amortised faster.

This process goes on until that good is no longer produced in the high wage country. Then the price of the good starts falling and some other good becomes equally profitable to produce in the low wage country and the process is repeated with that good. By going from good to good the low wage country industrialises and eventually has full employment and, perhaps, rising wages. If the high wage country continues investing in making the capital goods used by the low wage country and in capital goods to make them, or in the capital goods to make the consumption goods to meet the increased demand from the income increases because of wage payments and profit, it can continue to keep its high wage and full employment.

Trade does not have to balance. B's exports are assured by competition and its imports by the desire to invest and to consume out of its increasing income from cloth production. Whether they balance or not is an assumption and it may be convenient to assume they do. But it is assumed for the present that capital is not mobile. If it were, all investment would be attracted by the higher profit margins to the low wage country until the wage rates in both countries became equal.

An alternative is that the high wage country does not adjust to trade in this coordinated way, but protects its domestic production. For simplicity the protection is assumed to be by *ad valorem* tariffs. While the low wage country production is displacing the imports of a good it obtains the price of the high wage country, but the price it obtains when it exports is lower by the tariff, the difference accruing to the high wage country as government revenue.¹¹ Its profit margin and income from production are lowered.

¹¹ Gresser makes this point for Bangladesh. Gresser, "America's Hidden Tax on the Poor. The Case for Reforming U.S. Tariff Policy."

Instead of progressing from good to good by displacing all the high wage country production of those goods, it may progress from displacing one import after another, depending on how high the tariffs on the different goods are. If the tariffs are high enough on all the exports, the profit and income of the low wage country from substituting for imports fall as production proceeds from the most profitable to the least profitable and are lower still when the country starts to export.

But there is a lower bound to the low wage country's profit rate. First, assuming that a good being produced is a consumption good, the profit margin on its production at the prices of the high wage country was higher than would have been the margin from making the capital good with which it was produced. Then, if the high wage country tariff is raised high enough, the profit margin is lowered enough for it to be equally profitable to make the capital good in the low wage country and use it there or, if there is no tariff on it, perhaps to export it. Now, to make that capital good, the low wage country imports another capital good, for which it must still export something and, again, if the tariff on that export is raised high enough it becomes more profitable to produce this other capital good in the low wage country, and so on. At some point the low wage country becomes autarkic and produces all the capital equipment it needs once it has installed the basic equipment for producing the rest. It can do this with positive profit margins, for not being able to do so would mean that the techniques of production available are not viable, whereas the existence of the high wage country shows they are. This is the common sense conclusion that the more the other countries protect their own production, the more a country should rely on itself.

A model serves both to show the consistency of the reasoning and to illustrate it. It also brings to notice phenomena that are not obvious and are complicated to describe purely verbally. It is a sequel to the model of Bensusan-Butt, though his was not designed to consider what happens when countries with low wages and little or no industry trade with industrialised countries with high wages, or what effects trade barriers, in particular tariffs, can have. In Bensusan-Butt's model capital goods can be made by labour alone, as well as by capital goods and labour together, and both countries begin devoid of capital goods. Hence, it is not suited for discussing how the prices of capital goods can be affected by tariffs on exports.

The model with one technique of production

There are two countries, A, the high wage country, and B, the low wage country, as described above, and, for simplicity, it is assumed for the time being that the highest profit margin that firms in B can obtain comes from making a consumption good, which will be called cloth, using machines imported from A.

At first it is assumed that there is only one technique of production for each good. Cloth is produced by a machine, to be called a b-machine, one man operating one b-machine producing one unit of cloth per period. b-machines are produced by k-machines, one man operating one k-machine producing one b-machine per period. k-machines can also produce k-machines, one man operating one k-machine producing μ k-machines per period.

Denoting the nominal wage in country A by W and the rate of return by R and assuming that machines last forever:

$$W + R.B = P$$

$$W + R.K = B$$

$$W + R.K = \mu.K,$$

where P , B and K are the prices of cloth, b-machines and k-machines respectively.

In the case that the highest rate of return obtainable in country B is obtained from producing cloth with imported b-machines, given the prices and denoting the nominal wage in country B by w and the rate of return by r :

$$w + r.B = P$$

$$w + r.K > B$$

$$w + r.K > \mu.K.$$

Then, denoting by r' the rate of return that country B would get if it imported k-machines, made b-machines and with these made cloth,

$$w + r'.(w + r'.K) = P. \quad r > r'.$$

Then: $w + r.(w + r.K) > P,$

which is another way to say that country B gets a lower rate of return if it makes its own b-machines.

If country A applies an *ad valorem* tariff, t , on imports of cloth, the price country B gets for cloth is $P/(1 + t)$.

Then: $w + r.B = P/(1 + t)$

$$w + r'.(w + r'.K) = P/(1 + t).$$

But, $w + R.K < W + R.K = B$. That is, since country B has a lower wage, b-machines are cheaper to produce in B at the same rate of return as in A. So, if A's tariff is raised so high that:

$$P/(1+t) = w + R.B, \quad \text{that is, such that } r = R,$$

$$\text{then } w + R.(w + R.K) < P/(1+t).$$

At this tariff, $r' > R$. Then there is a $t' > 0$ such that, $r' > r$ if $t > t'$. This means that, if tariffs are raised high enough, country B obtains a higher return from producing its own machines for making cloth. Moreover, if country B makes its own k-machines, it is assured a positive rate of return, r'' , given by:

$$w + r''.K = \mu.K$$

and this occurs for $t < P/w - 1$, that is at a level for country A's tariff that still allows a positive rate of return on country B's exports.

This is the common sense conclusion mentioned above, that the greater the protection by the high wage countries, the more self-reliant the low wage countries ought to be.

Several techniques of production

The model can be elaborated to include choice of technique and depreciation. The consumption good, cloth, remains the same, but b-machines are of various types, the type being denoted by a suffix. A b_y -machine is manned by one worker and has an output of α_y of units of cloth per period. The number of b_y -machines produced by a k_x -machine and a worker is β_{xy} . For simplicity, rather than having various types of k-machines make various types of k-machines, k-machines are assumed to be made by m-machines, of which there is only one type. An m-machine is manned by one worker and the number of k_x -machines it can produce per period is κ_x and the number of m-machines it can produce instead is μ .

At the start A makes all the machines it needs to make cloth and to replace machines as they wear out. Cloth is made with b_i -machines, which are made by k_s -machines. Any other combination results either in a higher price of cloth relative to W or a lower R . Each machine is assumed to last T periods and then to collapse, when it becomes valueless and is disposed of costlessly. The prices in A are given by the following equations:

$$W + (1 + R).M_1 = \mu.M_1 + M_2,$$

$$W + (1 + R).M_p = \mu.M_1 + M_{p+1},$$

$$W + (1 + R).M_T = \mu.M_1,$$

$$W + (1 + R).M_1 = \kappa_s.K_{s1} + M_2,$$

$$W + (1 + R).M_p = \kappa_s.K_{s1} + M_{p+1}$$

$$W + (1 + R).M_T = \kappa_s.K_{s1},$$

$$W + (1 + R).K_{s1} = \beta_{st}.B_{t1} + K_{s2},$$

$$W + (1 + R).K_{sp} = \beta_{st}.B_{t1} + K_{sp+1},$$

$$W + (1 + R).K_{sT} = \beta_{st}.B_{t1},$$

$$W + (1 + R).B_{t1} = \alpha_t.P + B_{t2},$$

$$W + (1 + R).B_{tp} = \alpha_t.P + B_{tp+1},$$

$$W + (1 + R).B_{tT} = \alpha_t.P.$$

Here B_{tp} , K_{sp} and M_p denote the prices of the corresponding machines in their p^{th} periods and P is the price of cloth. These equations also imply that firms are indifferent to using an m -machine of a given age for producing k -machines or m -machines.

Omitting the age suffix, the equations for new machines and cloth can be summarised as:

$$1.a \quad \mu.M = W + h(R).M,$$

$$1.b \quad \kappa_s.K_s = W + h(R).M,$$

$$2.a \quad \beta_{st}.B_t = W + h(R).K_s,$$

$$2.b \quad \alpha_t.P = W + h(R).B_t;$$

where $h(R) = R.(1+R)^T / [(1+R)^T - 1]$, which is monotonically increasing if $R > 0$.

When firms in B start to invest in machines they choose those that yield the highest return. For the present the prices are taken as given, the effects of B's investment and production on them are taken up later. This avoids complications that obscure the argument without changing the main conclusions.

Several techniques of production: free trade

The highest rate of return obtainable in B is assumed to come from importing b_x -machines made in A with k_v -machines and producing cloth. Then:

$$\kappa_v.K_v = W + h(R).M$$

$$3.a \quad \beta_{vx}.B_x = W + h(R).K_v,$$

$$3.b \quad \alpha_x.P = w + h(r).B_x.$$

Since $w < W$, $r > R$.

By hypothesis, if b_t -machines and b_x -machines are different, making b_t -machines in B and making them in A but operating them in B yield lower returns. For the same reason, b_x -machines are not operated in A. Then:

$$4.a \quad w + h(r).K_s > \beta_{st}.B_t$$

$$4.b \quad w + h(r).B_t > \alpha_t.P$$

$$5.a \quad w + h(r).K_v > \beta_{vx}.B_x$$

$$5.b \quad W + h(R).B_x > \alpha_x.P$$

From these follow: $B_x < B_t$, K_s , K_v , which is to say that, because its wage is lower, country B uses less capital per head than A and has a lower cloth output per head.

If, instead, the highest rate of return obtainable in B were to come from importing k_y -machines and making and operating b_z -machines there, B_z and r would be defined by:

$$6.a \quad w + h(r).K_y = \beta_{yz}.B_z$$

$$6.b \quad w + h(r).B_z = \alpha_z.P.$$

3.a is replaced by 6.a and 6.b has the same form as 3.b. By hypothesis, b_z -machines cost more to produce or to operate in A, so:

$$7.a \quad W + h(R).K_y > \beta_{yz}.B_z$$

$$7.b \quad W + h(R).B_z > \alpha_z.P.$$

Then B_z , $K_y < B_t$, K_s and $\alpha_z < \alpha_t$. Again, B uses less capital per head and has a lower output per head than A. B_z is a tradable good that is made and used, but cannot have a world price.

At some moment firms in B begin to invest and they choose the activity that gives the highest rate of return. In the following discussion of industrialisation in B the highest rate of return obtainable in B in free trade is assumed to be given by importing b_x -machines made in A with k_v -machines and producing cloth.

The effect of the growth of output on prices can be discussed in two ways. One is to assume that firms in both countries act with perfect foresight and that they and workers move freely from one activity to another. Then, as B's output of cloth increases, A's firms withdraw from the production of cloth. If the growth of B's output is so fast that some b_t -machines in A are scrapped before they cease to be usable, these machines are amortised faster to yield the same rate of return. Similarly, k_s -machines are scrapped early and amortised faster, to the extent that they are not transferred to producing b_x -machines, or even other b -machines. Hence, A's firms reduce their own production of cloth in such a manner as to allow prices to rise by enough to obtain the same rates of return as before on investments that last a shorter time. It is possible that B's demand for b -machines does not create enough

employment in their production in A to offset the loss of jobs in the production of cloth, but it is also possible that the demand is too great to be satisfied without raising prices and profit rates in A. It may also be necessary to increase the stock of m-machines to produce the k-machines needed to meet the demand for b-machines. In each case the price of cloth rises and perhaps so does the price of b_x-machines. A's real wage falls, though less the slower is B's growth. If the growing employment and profit in B result in greater value of capital per head in A, it has been preceded by more saving and the total profit in A is higher.

This is equilibrium in the sense that firms' expectations are satisfied. But explicit formulae for prices become unmanageably complicated, if not impossible. They depend on the rate of investment in B and the relative sizes of A's and B's cloth producing sectors, and then the possibility arises that, because of the price changes, other types of machines become more profitable in either country. These are complications that Bensusan-Butt avoids by assuming, that machines last forever, can be used at any time to produce any good and are only of one type. But the effects of amortisation, the dependence of the length of use of an investment on growth and the choice of techniques can be discussed satisfactorily in qualitative terms and explicit price formulae are unnecessary. Equations 1-4 are adequate reference points for the discussion of prices and rates of return that follows, and will, for the most part, be used as such.

The second way to discuss the growth of B's output of cloth is to assume that A's firms try to compete to keep their markets for cloth and are forced to cut their production because they can always be undercut. Cloth prices can vary anywhere between the level that just covers the wage cost in A and that, discussed above, given by perfect foresight, and the returns are accordingly below those obtained with perfect foresight. In A employment in the production of cloth and of b-machines for use in A falls and unemployment can occur to the extent that B's demand for b-machines does not create enough jobs. Real income in B does not rise as much as with perfect foresight, so that growth is slower for the same saving rate, and real income from profit in A falls, though the real wage there rises.

In either case a point comes when the growth of cloth output in B causes the price to fall enough that some other good becomes more profitable to produce there. Other consumption goods can be assumed to exist and to be made with similar hierarchies of machines, though the next most profitable good could equally be a capital good. This progression from one good to the next can continue until there is no more unemployment in B, the profit rate there declining all the while but real wages rising in both A and B. Nominal wages in B can be expected to rise at this stage, though there is no

necessity for it. If they do rise, the profit rate in B falls more. If they do not, a stationary state follows.

In principle, during this progression, if trade balances, any unemployment in A is minor. At first, B's demand for b-machines and consumer goods other than cloth grows as income from profit increases and is met from production in A. If B can have trade deficits, there may be excess demand for A's products and no unemployment. All that is required is that A's firms and workers adapt to change and production shift accordingly. Later, as other goods are produced in B, A supplies the machines for making them and the consumption goods not produced in B.

Trade and development with tariffs

When the high wage countries have trade barriers that lower the prices received by the low wage country firms, the profit rates in the latter are lowered and, if taken far enough, this makes the production of capital goods in the low wage country more profitable than production of cloth with imported machines. This can be seen with an *ad valorem* tariff. If the price of cloth in country A is P and the tariff is t, the price that country B receives for the cloth it exports to A is $P/(1+t)$, the price difference accruing to A as government revenue. Assuming that prices in A are given by equations 1.a to 2.b, the rate of return to producing cloth in B using b_x-machines is given by:

$$w + h(r).B_x = \alpha_x.P/(1+t).$$

If t is put so high that the price of cloth only covers the wage cost, $w = \alpha_x.P/(1+t)$, country B can only obtain a positive return from importing b-machines by choosing types of these machines that have higher outputs. The higher the tariff, the higher the output of country B's b-machines must be and, therefore, the more capital intensive its cloth production. But, if t is so high that B cannot obtain a positive return using the same b-machines as A, $w = \alpha_x.P/(1+t)$, it can obtain a positive return by importing k_s-machines and making b_t-machines for export to A, because, if r is determined by

$$w + h(r).K_s = \beta_{st}.B_t = W + h(R).K_s, \quad w < W \text{ implies } r > R.$$

Then: $W + [h(R)/\beta_{st}].[w + h(r).K_s] = \alpha_t.P.$

This value of r is assured given this level of t, but a higher value may be possible with a suitable choice of types of b-machines and k-machines, the highest value being given by:

$$8. \quad \text{Max}_{mn} \left\{ r \mid W + [h(R)/\beta_{mn}].[w + h(r).K_m] = \alpha_n.P \right\}$$

For some lower values of t , B may obtain its highest rate of return from importing k -machines and making b -machines and using these to make cloth domestically. For, if t is such that B's rate of return on using imported b_x -machines to make cloth is reduced to the same rate of return, R , as in A,

$$\alpha_x.P/(1+t) = w + h(R).B_x = w + [h(R)/\beta_{vx}].[W + h(R).K_v] > w + [h(R)/\beta_{vx}].[w + h(R).K_v],$$

which says that if B, at the same return as in A, imports k_v -machines and makes b_x -machines and uses these to make cloth, its price will be below $P/(1+t)$, i.e. it can obtain a higher return. Again, this return is assured, but, by choosing the types of k -machines and b -machines suitably, it may get a higher rate still:

$$9. \quad \text{Max}_{yz} \left\{ r \mid w + [h(r)/\beta_{yz}].[w + h(r).K_y] = \alpha_z.P/(1+t) \right\}$$

and $r > R$.

Since the expression 8 is not a function of t and since r in the expression 9 increases as t decreases, there are values t' and t'' such that the highest value of r obtainable is given by 8 for $t \geq t''$ and by 9 for $t'' \geq t \geq t'$. For $t \leq t'$, B's highest rate of return is obtained from importing b -machines and making cloth.

Then, if $t \geq t''$, B exports b -machines to A, displacing A's production of b -machines and the same sequence follows as was described for cloth under free trade and, in the same way, it either ends when A produces no more b -machines, whereupon B starts exporting some other good, or when A imposes a high enough tariff on imports of b -machines. In the latter case B obtains m -machines from A and exports k -machines and, if A puts a high enough tariff on imports of k -machines, B becomes autarkic, except for the imports in exchange for its corn, which puts a lower limit to the rate of profit in B.

And, if $t'' \geq t \geq t'$, B exports cloth to A, but imports k -machines and makes its own b -machines. As B's exports of cloth grow and continue to displace A's cloth production, A may raise t until $t \geq t''$, whereupon B starts exporting b -machines, as just described.

It is possible that all cloth production is transferred to B, at which point the price of cloth is no longer determined by A. If B still has unemployed workers and the nominal wage does not change, the price of cloth falls to yield the same rate of return as the production there of b -machines. If A's tariffs lower the prices of b -machines, the price of cloth falls accordingly and so on for tariffs that lower the prices of k -machines.

3. THE FACTOR ENDOWMENTS THEORY: WHAT IS A FACTOR?

A model of international trade cannot have assumptions closer to those of the factor endowments theory than those made here and also have durable, heterogeneous capital goods that are all produced with capital goods. The model here is as simple as it can be and that it differs from the factor endowments theory in beginning with assumptions about nominal wage rates is the consequence of having tradable produced factors. When factors are produced and traded, each is produced where it costs least, and that requires that the costs of production, hence nominal wages, of all countries be known first. There is a dual to this; a factor is installed in a country because of the return it yields, which depends on the costs of production and prices of the goods and, therefore, on wages. Endowments at any time of produced factors are, therefore, the consequences of, among other things, the nominal wage rates of the various countries and cannot be said to determine them. Besides, as already pointed out, wage differences are a permanent fact that must have consequences and these must be considered. It cannot be objected, either, that the starting conditions given here fall outside the scope of the factor endowments theory, which has its own specific initial conditions for validity. That would be to eliminate it as a theory by confining it to identical or, at least, similar economies.

To the extent that the model here is incompatible with the factor endowments theory, therefore, the fault cannot lie with the model. But there are many reasons for finding fault with the factor endowments theory. This is seen in the following by examining the assumptions of the theory and the cogency of the reasoning from those assumptions, by comparing the theory with reality and by explaining the differences.

Same production functions

Both Heckscher and Ohlin made the assumption that production functions were the same for all countries, though neither believed he was making an assumption. In his exposition of the theory Ohlin states, "The supply of goods ... depends ultimately on ... the supplies of productive factors, and ... the physical conditions of production. These conditions – the natural and unchanging properties of the physical world which are everywhere the same – determine ... the technical process, with due consideration of their prices ..." ¹². In other words, production functions must be the same because they represent the 'properties of the physical world which are everywhere the same'. He repeats the point, "Differences in technique are clearly unthinkable if relative prices of *all* factors are the same; for the proportions in which the productive factors are combined – the

¹² Ohlin, *Interregional and International Trade*, 14.

technical coefficients – are functions of relative factor prices. The quality of factors being the same, which is so far assumed, *the forms of these functions* must also be identical in the two countries. If relative factor prices coincide, the technical process must be alike in A and B.”¹³ In the mathematical appendix to his book (Appendix 1) the production functions of the two countries are the same.

It is an assumption, nonetheless. The techniques of production are not directly derived from the ‘properties of the physical world’, but from what is known of them. To take an example, a plant that thrives in one place may not do so well in another that appears to have the same conditions. Research can be expected to discover the reasons and may find means to enable the plant to grow equally well in both places, but this is the result of an acquisition of knowledge. As knowledge accumulates, the techniques of production improve, though the properties of the physical world stay the same. Were Ohlin’s argument for the uniformity of production functions to be taken literally, technical progress would be ruled out and so would such conspicuous features of modern economies as proprietary knowledge and branding of products.

Nevertheless, it is assumed here that, regarded as purely physical processes, the techniques of production in one country operate much the same as in another. That they derive from what is known of the physical world does not mean that they vary much from country to country. Consequently, the manufacturing techniques for steel, motor cars, light bulbs, cloth and other goods used in one country will work the same way in other countries, with modifications, which can be taken as minor, to allow for environmental differences, such as climate. The possibility that a manufacturing process used in one country might not work in another or work differently cannot logically be ruled out; since the laws of physics are not fully understood and complex systems are often not calculable or predictable, unknown effects that prevent a manufacturing process from working in one place as in another cannot be excluded *a priori*. But they can be assumed away in practice. The same is true for agriculture, where the possibilities of countries differ markedly – apples are not grown in the Sahara, nor coffee in the wheat lands of the US or cinnamon in Norway. Here the differences can be attributed to the availability of factors, combinations of soil and environmental features, rather than to differences in the way vegetal processes work. In what follows, the same techniques of production will be assumed available to all countries, unless otherwise stated.

¹³ Ohlin, 15.

The determination of factors

A separate problem is that there is no criterion by which factors can be identified. Some of the initial appeal of the theory came from the appearance of simplicity; it seemed that all that was needed to apply the theory was to identify and quantify factors and determine patterns of demand. Factors and consumption being apparently directly observable, the only problems to be expected in testing the theory and using it to study international trade would be the practical ones of getting data. But there are conceptual problems, apart from that of produced factors already discussed.

Discussion of what makes a factor a factor is rare, though it should be part of the explanation of international trade using the factor endowments theory; with few exceptions, international trade theorists take factors for granted. They may not specify them, or even their number, to avoid limiting unnecessarily the scope of their arguments, but they regard them as given. Questions may arise as to whether or not the inclusion of a neglected factor could alter the outcome of some theoretical or empirical work, as happened, for instance, with the addition of the factor, land or natural resources, to the factors capital and labour to resolve Leontief's paradox, but the factors are considered well defined. Ohlin, as the first expositor of the factor endowments theory to a wide public, was one of the few to discuss what he meant by a factor. Others who tried to explain what they meant, Ivor Pearce, for example, differed little from Ohlin in substance. Consequently, the discussion that follows refers principally to Ohlin.

The first question is, how much freedom is there in the choice of factors? Can factors be chosen according to the problem at hand, or are they independently determined so that every discussion of international trade must necessarily take all of them into account? Ohlin epitomises the former, flexible view and all those who have accepted the factor endowments theory seem to have followed him in this.

He was at pains to explain what he meant by a factor, but, far from providing a criterion, he insisted that the choice was to be made according to the problem at hand. Starting with labour, "There are cases where ... no finer classification is necessary ...". But, "... if differences in wages between groups of workers in a country last for a sufficient period of time, and influence the nature of the international division of labour, then these groups may well be regarded as separate productive factors ..."¹⁴ "What division into separate labour factors is to be used must depend on the nature of the special problem under discussion." "By factor of production we shall mean, ...so far as labour is concerned, a group that for some period of time varying

¹⁴ Ohlin, 69.

with the nature of the analysis, is fairly non-competing with other groups."¹⁵ "In most cases a broad division into three factors only will suffice: (1) 'unskilled labour', (2) 'skilled labour', and (3) 'technical labour'."¹⁶ Since differences between groups of unskilled labour can last a long time, "it is convenient and fruitful to regard them as separate *sub-factors* ..."¹⁷ The same applies to skilled and technical labour. International comparisons of labour increase the diversity; for instance, Swedes and Estonians, Englishmen and Italians "... may properly be regarded as different sub-factors."¹⁸ It is necessary sometimes even to distinguish between Danes and Swedes or between Englishmen and Frenchmen.¹⁹

Natural resources are to be treated in the same way. Ohlin identifies five qualities to group natural resources²⁰, but adds, "In almost all cases ... it is of paramount importance to reckon with a much greater number of factors.... The possibilities of qualitative differences with respect to soil, climate, wind, humidity, surface, etc., are immense. It is practical, therefore, further to divide each factor into sub-factors."²¹

This raises two problems. One is that, a specific factor having been defined, it must be shown that it behaves in the way prescribed by the theory. In the natural sciences and mathematics, the procedure when an abstract entity is defined is to demonstrate its properties. Ohlin neglects to do this. An illustration is Ohlin's definition of capital as a factor: "For the purposes of comparison, the capital available in a country is expressed as a sum of money that represents the cost of reproducing the capital goods in existence after deduction for depreciation and obsolescence."²² It is an abstraction: "To avoid misunderstanding it should be added that the expression 'mobility of capital' refers to abstract capital, not capital goods."²³ It seems there are two kinds of capital, one that determines trade and another that is mobile. The first is a quantity with a price:²⁴ "The price of this capital during a period of time is the rate of interest."²⁵ This, capital is well defined; its quantity is simply the book value of capital goods in use. But Ohlin does not show why capital so defined should behave like other factors. If, for instance, the

¹⁵ Ohlin, 70.

¹⁶ Ohlin, 71.

¹⁷ Ohlin, 73.

¹⁸ Ohlin, 80.

¹⁹ Ohlin, 80.

²⁰ "(1) Agriculture and forest growing, (2) fishing and hunting, (3) the production of minerals, (4) the production of water-power, and (5) transport activities." Ohlin, 76.

²¹ Ohlin, 76.

²² Ohlin, 54.

²³ Ohlin, 77.

²⁴ Ohlin, 54.

²⁵ Ohlin, 54.

pattern of trade were to change and a country with relatively much capital changed from making one set of manufactures to making another and its firms had to change their capital equipment accordingly, the book value of the old equipment would be irrelevant to the cost of the new equipment that would have to be manufactured as a new investment.

The second problem is that Ohlin's division of factors into more factors and sub-factors according to the needs of the argument is not possible without excessive restrictions. Ohlin believed, though he did not say it explicitly, that the theory could proceed in stages from the general to the particular; a country with a relative abundance of labour would export goods whose production was labour intensive and, other things being equal, the types of these goods would be determined by the relative amounts of various types of labour. The relative abundance or scarcity of broadly defined factors determined the broad categories of goods that would be exported and, the more finely factors were sub-divided, the more detailed the explanation of exports.

The fallacy can be seen in purely logical terms by considering the reverse process of combining different factors into a single composite factor and different goods into a single composite good. For instance, if two factors, factor 1 and factor 2, are used in the production of two goods, good 1 and good 2, a composite factor would be a function of both factors and similarly for the composite good. The output of the composite good would, then, be given by a production function with the composite factor as an argument, but not factor 1 or factor 2 separately. Then, variations in the amounts of factor 1 and factor 2 that do not change the amount of the composite factor must leave the output of the composite good unchanged. The composite factor would be given by a function whose two arguments would be the two sums of the amounts of factor 1 and factor 2 used in the production of the two goods, i.e. of four given quantities. Any variation of any three of these quantities, within limits, can be compensated for by a variation of the fourth so as to keep the amount of the composite factor constant. This leaves enough room for several variations in each of which the ratio of the change in the output of good 1 to the output of good 2 is different and consequently the amount of the composite good is changed without change in the amount of the composite factor. In the special case that this does not happen for any change in the use of factor 1 and factor 2 that keeps the amount of the composite factor constant, the production functions of the two goods and the composition functions of the goods and factors must be related.

Stated mathematically, Ohlin asserts that a factor can be a function of several sub-factors and that variations in the amounts of the sub-factors that do not change the quantity of the factor make no difference to the overall production possibilities. This is to say that it makes no difference whether

the production possibilities are written as functions of the factors or of their sub-factors, which Leontief showed in 1947 implies that the ratio of the derivatives with respect to the sub-factors of the function representing the production possibilities be independent of all the other factors and their sub-factors.²⁶ In the example given here, it can be shown that the production functions of good 1 and good 2 are not independent.

That such progressive decomposition of factors and goods is not normally possible can also be seen less formally by considering two countries with two factors, land and labour, in different proportions. One country will then export land intensive goods and import labour intensive goods. If land intensive goods are decomposed into two types, crops and minerals, and land into two types of land, it is possible that the economy with the higher ratio of labour to land exports one type of land intensive good. This can occur if crops use one type of land, cropland, intensively and little or none of the other type, mines, whereas the reverse holds for minerals, and if a high proportion of one country's land is mines and a high proportion of the other's is cropland. In this case the use of a composite factor, land, and of a composite good was wrong from the start since the crops and minerals are traded in different directions. But, limiting decomposition to goods whose component goods are not traded in different directions and to factors that permit production functions specific to the goods may not be possible either.

Since the progressive decomposition of goods and factors is not in general possible, flexibility in choosing them is ruled out; goods and factors must be specified. Ohlin's method of drawing conclusions from broadly defined goods and factors and proceeding to greater detail is inconsistent with his theory. This does not logically preclude models of trade in which the factors change from period to period, though such models, if they exist, would be expected to include the causes of the changes and descriptions of the transitions between periods.

Theoretical arguments can omit specifying factors when carried out in the abstract, but not when drawing conclusions about the real world. As a rule, factors in theoretical models of international trade are taken as given and not progressively divided into more factors and sub-factors in the manner of Ohlin. But, in applications of these models to the real world, the factors must either be specified or the conclusions be shown to be independent of their number and nature. In the latter case the conclusions can presumably be derived without recourse to the theory. Going as far as an empirical test of some application of the factor endowments theory, or of the theory itself, the factors would have to be specified. An example is given by Leontief's

²⁶ Leontief, "Introduction to a Theory of the Internal Structure of Functional Relationships."

“paradox”.²⁷ When Leontief calculated that, apparently contrary to the theory, the US’s exports were less capital intensive than its import competing products, one of the arguments to save the theory was that he had considered only two factors and should have included a third, land. Since the US also has much land, both the relative abundance of factors and the data used in the calculations would be altered. Whether this resolves the paradox or not, it illustrates how empirical work that omits factors, at least factors that may be important, is open to question.

An attempt to use the factor endowments theory to explain international trade should, then, begin with a list of factors. Ideally, there would be an authoritative list of, at least, the important factors that determine world trade, a list compiled according to an objective criterion and accepted by all economists. The non-economist who learns that international trade is determined by each country’s endowments of factors and then asks what these factors are would wonder that no such list exists. He would wonder more to learn that the reason is not that international trade theorists do not agree on one, but that they appear not to deem one necessary and have not tried to establish one.

Relativity of factor endowments

Ohlin expounded his theory in terms of relative endowments and intensities of use of factors without apparently considering that the relativity might not be well defined. If the factors are more than two, a factor that is abundant relative to one factor in one country may be abundant relative to another in the other country. Comparisons of relative abundance or relative intensity require that it be specified relative to what. Ohlin never does that, though his penchant for specifying or defining factors according to the circumstances should have drawn attention to the need.

Relativity is well defined in the special case of two factors; it can be defined in terms of either factor since the other merely gives the inverse. Part of the popularity of the two factor, two good, two country model is explained by this. Since relative abundance is unambiguous, the direction of trade is normally determined by endowments, independently of preferences. Then each country exports the good that uses more intensively the factor with which the country is relatively well endowed, provided preferences do not differ too much between countries and the relative intensity of factor use of the two goods is not reversed.

An unambiguous definition of relative abundance of factors when there are more than two was given by Vanek. It is obtained by ranking a country’s

²⁷ Leontief, “Domestic Production and Foreign Trade.”

factors according to the country's share of each factor in the world total of that factor. One factor is, then, relatively more abundant in a country than another if its ranking is higher. Vanek used this criterion to argue that, assuming all countries to have the same factor prices, the goods of which a country is a net exporter will together use more of the factors with which the country is relatively well endowed and those of which the country is a net importer will use more of the factors of which it has relatively little. Put differently, each country exports the factors with which it is relatively well endowed and conversely for factors that are scarcer, the point separating the two sets of factors depending on the specific details. Since there is no equivalent method for ranking the factor intensities of goods, the actual composition of trade, which goods and in what amounts, cannot be specified and any country may export and import the same good.

4. THE FACTOR ENDOWMENTS THEORY: GENERAL EQUILIBRIUM AND PRODUCED FACTORS

The theorems of single period models

Without any attempt to establish what the factors are that determine international trade and how each country is endowed with them, the theory becomes models in the abstract. They are also single period models, meaning that all quantities are determined simultaneously and there are no earlier or later periods, which also requires balanced trade for there to be an equilibrium. When they are models of comparative advantage in the sense used here, the prices of goods and factors depend on the trade balance since that determines production in each country.

Since all countries are assumed to have access to the same production techniques the reasons for trade are differences of the proportions of the various factors with which countries are endowed and differences of preferences, which latter may be differences between countries or differences between the various classes of owners of factors. Samuelson provided, in 1953, a general mathematical formulation of the factor endowments theory with many goods and many factors, though with no produced inputs or joint production; each good has a production function with constant returns to scale and diminishing returns for individual factors.²⁸ It is taken here as given that an equilibrium exists in which the trade of every country balances; Arrow and Hahn give a proof for a general equilibrium model of this sort using a fixed point theorem²⁹ and Chipman assumes an equilibrium of the same kind in his survey of international trade theory.³⁰

²⁸ Samuelson, "Prices of Factors and Good in General Equilibrium."

²⁹ Arrow and Hahn, *General competitive analysis*.

³⁰ Chipman, "A Survey of the Theory of International Trade."

Models of the type described by Samuelson with only one primary factor give no reason for trade. All countries have access to the same production techniques, so the factor's price must be the same everywhere for trade to balance and, since there is only one income class, differences of preferences do not give reason for trade. Ricardo's reason for trade with one factor was differences in the production techniques.

The simplest factor endowments model that gives a reason for trade has been that of two goods produced with two factors. From it follow three "theorems", the Rybczynski theorem, that an "increase in any factor endowment, at constant terms of trade, will cause a greater than proportionate increase in the output of the good intensive to that factor and a decline in the output of the other good",³¹ second, the Stolper-Samuelson theorem, that, in a country with given quantities of factors a higher price for one good relative to the other results in the price, relative to the prices of either good, of the factor used more intensively in the production of the first good being higher and that of the other factor being lower, and, third, the factor price equalisation theorem, which states that factor prices of different countries are equal provided the factor endowments of the countries are not too different and that the factor either production function uses more intensively is the same for all factor prices, i.e. there are no factor intensity reversals.

Much of the development of the factor endowments theory has consisted of finding under what conditions the theorems hold, perhaps in modified form, when the model is made more general or otherwise modified. The first two theorems refer to the difference made by changes of some quantities, not to change over time, and the Rybczynski and Stolper-Samuelson theorems are assertions about alternatives that are not directly observable. In contrast factor price equalisation is an assertion about the given situation and is, in principle, observable and can, therefore, be compared to reality.

Extending the model to more goods and factors in the abstract amounts to a set of mathematical propositions about the relations between their prices and quantities, in particular to propositions about the conditions for the three theorems to hold. One extension is to have more goods and factors and another is to include goods as inputs. Joint production is always excluded, although models with joint production have been devised but, having fixed input-output coefficients, are not extensions and are not discussed here.³²

Samuelson, who first proved the factor price equalisation theorem, was also the first to try to generalise it to more than two goods. By using the implicit function theorem he showed that, if the numbers of factors and goods

³¹ Ethier, "The Theorems of International Trade in Time-Phased Economies," 232.

³² Uekawa, "Some Theorems of Trade with Joint Production."

are the same and if the Jacobian of the functions giving the prices of goods in terms of the prices of factors is non-singular, for given prices of the goods there are prices of the factors from which those goods prices follow and that they are determined uniquely for all goods prices, provided they do not differ too much from the given prices. Conditions on the Jacobian can be found sufficient to remove the proviso. Similarly, the Rybczynski and Stolper-Samuelson theorems hold, along with some variations, given the appropriate Jacobians and production functions. These developments of the factor endowments theory consist of showing how the properties of mathematical entities, like the principal minors of determinants, positive matrices, Minkowski matrices and so on can affect trade.

Produced inputs and profit margins

At issue here is whether or not the factor endowments theory can be adapted to accommodate produced factors. As was to be expected, the controversy of the 1960s over the aggregation of capital as a quantity prompted some economists to concern themselves with, the question of how capital could be a factor in the theory and, as Mainwaring put it, ‘... the consequences for some well-known international trade theorems of the inclusion of heterogeneous capital goods in the process of production.’³³

In several papers published in the 1970s Mainwaring, Metcalfe and Steedman argued that adding profit on produced inputs altered the conclusions from the standard factor endowments theory. Their procedure was to use single period models with two or more produced goods that were also inputs into production along with labour and, sometimes, other primary factors, so that the prices of goods were the sum of the costs of the primary inputs, the produced inputs and the profit on the produced inputs. The models varied in the numbers of primary factors, goods and techniques of production. In one case Steedman and Metcalfe seem to depart from the single period by assuming a steady state with unchanging prices, though, since they assume no durable capital goods and returns to scale are constant, it amounts to no more than saying that only the proportions of goods and factors matter, but is, otherwise, just a single period model.³⁴

Metcalfe and Steedman concluded from their model with two primary factors and two goods that, ‘... it is *not possible* to make any *a priori* predictions (*sic*) concerning the effect, on commodity prices and primary input prices, of differences in intensity of demand for commodities, or of differences in relative primary input supply.’³⁵ Even assuming that all

³³ Mainwaring, “Relative Prices and ‘Factor Price’ Equalisation in a Heterogeneous Capital Goods Model,” 109.

³⁴ Steedman and Metcalfe, “The Non-Substitution Theorem and International Trade Theory.”

³⁵ Metcalfe and Steedman, “Reswitching and Primary Input Use.”

countries have the same profit rate, ‘... nothing can be said *a priori* about the relationship between factor-prices and the factor-intensity of production methods, when the profit rate is positive, and it follows that nothing can be said *a priori* about the shape of the relative supply curve.’³⁶ Mainwaring concluded from his model, which had labour as the only primary factor, but many goods and techniques, that the relation between the relative prices of goods and the rate of profit is not monotonic and, given enough different techniques, that more than one rate of profit may result in the same set of relative prices. In each case, factor price equalisation does not hold.

Ethier defended the factor endowments theory using the same kind of model, with labour as the one primary factor, two countries, two goods, both of which are both final goods and inputs, production functions relating outputs to inputs of goods and labour, and a rate of profit on the value of capital, which is the value of the produced inputs. He reached the ‘simple and pronounced’ conclusion, that the ‘... assumption of two homogeneous timeless factors, and consequent neglect of the nature of capital, is quite harmless as far as the four basic theorems are concerned.’³⁷ His fourth theorem is the Heckscher-Ohlin theorem, which has a quantity version and a price version. In the quantity version each country exports the good the production of which is relatively intensive in its more abundant factor. In the price version the good each country exports is the one of which the price, without trade, was lower relative to the price of the other good than in the other country. Both factor price equalisation and this second version require the condition that there be no factor intensity reversal, that, in the production of each good, the factor used more intensively is the same through the relevant range of prices.

His argument for factor price equalisation does not bear out that the ‘neglect of the nature of capital’ is quite harmless, for it depends on production functions with the two goods and labour as inputs. This is no different to the aggregate capital and production functions, to which the objection has been that they are incompatible with heterogeneous capital goods. The two goods combine in different proportions to produce the same two goods; instead of the usual one malleable capital good, Ethier’s model has two. This allows the conclusion that, if there are no factor intensity reversals, the relation between factor prices and the relative prices of goods is monotonic, and factor price equalisation follows, whereas the correct procedure would have been to assume that different techniques of production use different capital goods, as in the model of Section 2.

³⁶ Steedman and Metcalfe, “Reswitching, Primary Inputs and the Heckscher-Ohlin-Samuelson Theory of Trade.”

³⁷ Ethier, “The Theorems of International Trade in Time-Phased Economies,” 226.

Since both countries use the same techniques of production, the country that produces proportionately more of the capital intensive good, i.e. the good with the greater value of produced inputs in its production per unit, has more capital per head. Ethier adds the assumption that countries have identical, homothetic preferences, i.e. goods are consumed in the same proportions, regardless of income or country, and obtains the quantity version of the Heckscher-Ohlin theorem. Without that assumption the assertion of the theorem is not true. It can be that the country with more capital per head exports the less capital intensive good because of the preference of its residents for the other good and a corresponding preference for its exported good in the other country. Alternatively, both countries may have the same preferences and this pattern of trade results from the distribution of income; the preferences of those whose income is profit are sufficiently toward the capital intensive good and the preferences of workers sufficiently toward the other.

The Rybczynski theorem is equally obvious. Taking the prices of goods as given, a greater amount of capital per head in the more capital abundant country implies a greater output of the corresponding good, which in turn implies less production of the other because the total amount of labour is fixed. Hence the output of the more capital intensive good is greater in proportion to the labour used in its production, whereas the overall amount of capital increases only in proportion to the difference between the capital per worker in the production of the two goods. If one good used no capital for its production, the increase of the total capital would be proportionate to the increased labour in the production of the other good. If both goods used the same amount of capital per head, there would be no change.

To prove the price version of the Heckscher-Ohlin theorem, Ethier defines 'holding real capital constant' when prices change. He does it for infinitesimal changes by specifying that any change in the value of the capital is wholly the effect of prices, which means that the effect of a change of production technique is exactly cancelled by the effect of change of output. Ethier proves, using the calculus, that if, say, the price of the more capital intensive good is infinitesimally higher, production techniques use less of it and, for real capital to be constant, the output of that good is higher. He does not explain how this applies to changes between having no trade and having trade, which are not infinitesimal. Ethier also uses the calculus to obtain the Stolper-Samuelson theorem.

Two of Ethier's theorems are only tautologies and not arguments in favour of the factor endowments theory. They state logical consequences of the assumptions, whereas the theory, valid or not, is a theory of cause and effect; it asserts that endowments and preferences result in specified results.

Otherwise Ethier's defence of the factor endowments theory depends on the assumption that is in dispute.

In these single period models what is referred to as capital is necessarily fully consumed in the period and is indistinguishable from raw materials. Normally capital goods are durable goods used over several periods and may also need more than one period for their making. Before Ethier, Kemp had already argued that the main conclusions of the factor endowments theory held also for durable capital goods.

Kemp used a model similar to Ethier's, with two goods, both inputs into the production of both goods, and two primary factors. As an input each good is a durable capital good and the price of each good is the sum of the costs of the factors and the profit on the value of the capital goods. As with Ethier, production is given by production functions. Kemp assumes a steady state, in which the rate of profit remains constant, factor supplies increase uniformly at the same rate through endless pasts and futures and outputs grow in fixed proportion. From this Kemp derives the Stolper-Samuelson theorem for two primary factors and both that and the Rybczynski theorem for just one primary factor with the rate of profit considered the other factor price.

Kemp argued that the theorems he proved are about comparisons of steady states, not of just a single periods, and his justification of steady state models was that the original Heckscher-Ohlin model is also of this kind.³⁸ Ethier asserts that his model can be interpreted as a stationary state, a steady state in which nothing changes, and allows "... intermediate goods to have any finite durabilities (*sic*) and any time patterns of productivity, and to be either purchased or rented."³⁹ He refers to Kemp and justifies the stationary state by stating that "... as Metcalfe and Steedman ... have pointed out, this is in any case appropriate for present purposes."⁴⁰ Metcalfe and Steedman were, however, using their model as a counterexample to Ethier's theorems and what was appropriate for their purpose was not necessarily appropriate for Ethier's purpose.

There are more objections to these models. First, like Ethier, Kemp's conclusions differ from those of Metcalfe and Steedman because he assumes production functions and malleable capital. Second, neither Ethier nor Kemp, nor any other proponent of such models, explains how the models can be used to depict international trade when countries cannot be assumed to grow at the same rates and the composition of trade must change with time. If there are primary factors other than labour, an explanation is also

³⁸ Kemp, "Heterogeneous Capital Goods and Long-Run Stolper-Samuelson Theorems," 253.

³⁹ Ethier, "The Theorems of International Trade in Time-Phased Economies," 226.

⁴⁰ Ethier, 226.

needed as to why they should all grow at the same rate. Third, Ethier's model cannot explain trade nor show why it has any benefits. If all countries have access to the same production techniques and have the same factor prices, they can produce the same goods at the same prices. They trade by assumption, but could produce the same composition of goods they consume with or without trade. If there is to be an argument for trade in such a model, it has to be that trade allows one country to have more capital per head and, therefore, more profit, which is in the interest of the owners of the capital goods in that country, but that is not a justification normally given for trade.

Intertemporal general equilibrium

If a model of international trade is to include produced means of production in general equilibrium, yet not be limited to steady states, it must be along the lines of the intertemporal general equilibrium model for the closed economy of Arrow and Debreu.⁴¹ This model's assumptions about the production and individual preferences seem to be as general as the conclusions allow, though it depends on a particular notion of human rationality, which, itself, depends on perfect foresight. It allows the proportions in which goods are produced and the rate of profit to change over time and capital goods are manufactures, not a factor. The model of international trade of Arrow and Hahn referred to above is not of this kind.

In its general form, the model has not been adapted to international trade, but, assuming for the sake of argument, that it can be done and an equilibrium exists when the same production techniques are available to all countries, if labour is the only primary factor and all goods are tradable, the whole world can be treated as a single economy. If primary factors, other than labour, are added and if they and some goods are not tradable, countries consist of blocks of individuals who have access to the same untradables. Further extensions would be to allow countries to use trade barriers and to let them differ in the technical possibilities to which they have access. Equilibrium can be assumed to exist for these models too.

Models like that of Arrow and Debreu are not to be taken as descriptions of closed economies, let alone of trade between open economies, though they are believed by some economists to help understand them. They are mathematical constructs and any assertion that they have some relevance to reality has to be substantiated with a statement of what that relevance is and an explanation of the mechanism by which the actual functioning of the economies brings it about. Arrow and Hahn limit the usefulness of the model to showing that it is logically possible that, 'a decentralized economy motivated by self-interest and guided by price signals would be compatible

⁴¹ The references here are to the exposition in Arrow and Hahn.

with a coherent disposition of economic resources that could be regarded, in a well-defined sense, as superior to a large class of possible alternative dispositions.⁴² They do not answer the questions people do put about how a private enterprise system functions. These questions are not answered by making assumptions about the behaviour of people and firms that do not approximate reality. A non-economist would be more likely to retort that the model refutes itself: if the assumptions have to be so far-fetched, there is cause for doubt.

One practical reason can be given for denying the usefulness of the model, namely that such intertemporal equilibrium eliminates balance of payments problems and, therefore, much of the matter of international economic theory. The balance of payments is not a separate constraint for, if individuals abide by their intertemporal budget constraints, countries do so too. Balance of payments problems are problems of disequilibrium and incompatible with general equilibrium. Adding balance of payments constraints may, perhaps, not alter the model's properties too much, but leaves the question as to what mechanism brings those constraints about.

5. THE FACTOR ENDOWMENTS THEORY: COMPARISON WITH REALITY

As reformulated by Arrow and Hahn and by Chipman the factor endowments theory lacks the concreteness and immediacy of the theory as propounded by Heckscher and Ohlin. Ohlin's many examples show that he and Heckscher believed that scarcity and abundance of factors were simple concepts and evident. In contrast, it is not apparent from the reformulated factor endowments theory what goods each country will produce and which it will export or import. Being a theory of comparative advantages, all parts are interdependent; a change in a factor or preferences in one place affects the outcome everywhere. In principle the theory would have to be represented by a mathematical model covering the whole world in which all goods and factors would have to be identified and production functions for all goods estimated for the whole range of observed factor prices. Like the model of Arrow and Debreu, it remains just an abstraction, except that it purports to describe the real world and should, therefore, be compared to reality.

Yet comparison with reality is not feasible; the model described is too complicated and cannot be simplified in a way that gives trustworthy results. Simplification by aggregating goods and factors, it has been seen, gives wrong results, but the alternative of omitting countries, factors and goods judged to have little effect on the whole cannot be presumed to be reliable without some independent way of knowing the margin of error, and there

⁴² Arrow and Hahn, *General competitive analysis*, vi–vii.

does not seem to be such a way. The more the model and the collection of data are simplified, the greater the error margin, and the less reason to believe the results.

Instead of using the theory to explain trade, the theory's proponents have tried to find evidence for the theory from examples of international trade or from inferences that can be tested. Ohlin believed that his many examples substantiated the theory. Inferences that can be tested must be such as can be made without knowledge of the general equilibrium, beyond some plausible restrictions. In such cases they can be expected to be qualitative. Apart from Ohlin's examples, there are three such inferences, each subject to conditions.

The first holds if the number of factors is two, capital and labour, though the numbers of goods and countries are not constrained. Then, if consumer preferences are not too different, the goods the country with the most capital per head exports require more capital per head to make than the goods that compete with imports. This was the proposition Leontief tested.

The second is factor price equalisation, which follows from the assumption that all countries have the same production functions and that the prices of goods are the same for all, provided that the countries produce at least as many of the same goods as there are factors. Being a conclusion from solving the equations relating factor and goods prices, it holds as long as the solution is unique, which requires that the factor endowments not be too different.

The third inference is the optimality of free trade. Unlike the first two, it holds for all types of comparative advantage and not only for the factor endowments theory. What distinguishes the factor endowments theory is that, since factors may be variously affected by trade, optimality in the sense given earlier is not the same as welfare and the possibility that policy can affect the income from some factors has to be examined.

Ohlin's Examples in Support of his Theory

Examples can be evidence for a theory if they are incompatible with some rival theory. That Kuwait exports oil is not evidence. An example must also avoid circular reasoning. Thus, it is often asserted that the US exports technology intensive goods because it is well endowed with "technology". That some of the goods that the US exports are technically advanced is not explained by supposing that "technology" is a factor in the sense of the theory, especially when it cannot be quantified as the theory requires. Here the reasoning is circular. A systematic account of examples that have been used as evidence for the factor endowments theory cannot be attempted

here. Rather, a few examples given by Ohlin will be considered, even though examples in textbooks nowadays are more up to date.

Ohlin seems not to have been aware that the flexibility he assumed he had in choosing factors and subdividing them reduced many of his examples to circular reasoning: when a country exports a certain good the factors used for producing that good can be identified and, so, the abundance of those factors explain the exports. Where institutions, proprietary knowledge and experience, technical progress and concentrations of producers are seen to determine exports, Ohlin finds factors and sub-factors. In principle he would have accepted counter-examples as evidence against the theory, but this procedure makes them virtually impossible to find since some factor or other can always be supposed.

The differences between Danish and Swedish farmers are held to account for the differences in their countries' agricultures, since the land and climate are much the same.⁴³ The development of the German chemical export industry before the First World War was the result of the supply of cheap intellectual labour of a certain quality⁴⁴. Northern Ireland and Southern Scotland manufactured, at the time of Ohlin's writing, the finest linens because they had acquired the skills earlier, when flax was grown in these areas (flax cultivation requires much labour and had moved to countries where wages were lower).⁴⁵ The characteristics of French technical, semi-skilled and unskilled labour explain the "... export from France of articles of luxury requiring a certain taste and handicraft ...".⁴⁶

Taking for granted that the differences between Danish and Swedish agriculture are, as Ohlin says, entirely due to differences between the farmers and not the result of differences in land and climate, why should such similar people be different factors? The question follows, would a Danish farmer transferred to a Swedish farm be different to a Swedish farmer? Ohlin answers it with the example of an Italian worker transferred to America, where he can be assumed to work just like any American worker. He concludes that it is better "... to classify all individuals as belonging to the same labour group, if under similar conditions as regards machinery and organisation they are found to be fairly equal in efficiency ...".⁴⁷ The Danish and Swedish farmers are not different after all.

The reason for the inconsistency is that the reasons why Danish and Swedish farmers differ cannot be reduced to factor endowments. Ohlin

⁴³ Ohlin, *Interregional and International Trade*, 81.

⁴⁴ Ohlin, 130. See also p.84.

⁴⁵ Ohlin, 136.

⁴⁶ Ohlin, 81.

⁴⁷ Ohlin, 80.

seems to believe it is because the Danes work in more cooperative forms of organisation than the Swedes. Whether this explanation is valid or not, if the farmers do differ, such an institutional component is needed to explain why. Differences in organisation can be a historical legacy or they can be the effects of population densities, of interaction with neighbouring countries and regions or the effects of other causes, but they are not factors and cannot usually be quantified. In other contexts factors may be created or caused by institutional differences, but not so in this context.

Ohlin's example of the German chemical industry before the First World War illustrates the inability of the theory to cope with technical progress. At that time the chemical industry, especially the German chemical industry, stood out for the rate at which it was developing new products and processes. Germany's success in exporting chemicals was in part the result of having new products and processes before its competitors, which, in turn, was made possible by the system that trained a large enough number of chemists to suitably high standards. In the same way, a large part of trade in manufactures and agricultural products nowadays is determined by the ability of individual companies to provide goods that are technically better than the products of competitors or have technical features that distinguish them. Even if part of the explanation of such trade is ample supplies of people able to do the necessary research and development and to supervise frequent changes of production processes, it falls outside the scope of the factor endowments theory because goods are continuously changing and it is the change that determines the direction of trade.

Historians would presumably accept Ohlin's assertion, that the products of the linen industry of Northern Ireland and Southern Scotland were exported because of their high quality in relation to price and that this quality was the result of experience gained over more than a century before 1933, when his book was published. From that assertion Ohlin argued that the workforce was a factor or sub-factor that explained the exports. His argument might have been valid if the workforce had been highly trained, like the German chemists. But linen manufacture was not technically especially difficult and the workers in these industries were not unusually skilled. An alternative argument is that the factors or sub-factors that distinguished the Northern Irish and Southern Scottish linen industries were not the workers on the factory floor, but the technical and managerial staff. It took longer to train technical staff than workers on the shop floor, which is why Ohlin holds these two groups to be different factors or sub-factors, but the experience was vested in them and manifested itself in the technical controls and organisation of the manufacturing. Nevertheless, just as with the Italian worker, individuals in the technical or managerial staffs of the linen industries

of different countries could be presumed to be the same factors because they could be assumed to be equally efficient under the same conditions.

What differentiated the technical staff of one country's linen industry from another's is that the conditions in the two countries' linen industries were not the same. There is other no reason for otherwise similar people to be of different levels of efficiency. To stress the obvious, the differences in skills and experience between the linen industries of Northern Ireland and southern Scotland, on the one hand, and of other countries, on the other, must have been transmitted to new workers who entered the workforces of these industries as old ones departed. In other words, the institutions in one country that store, develop and impart the knowledge and skill that make the industry in that country superior to its competitors in another, namely the enterprises themselves and the institutions serving the industry, differ from the corresponding institutions in the other country in having a better store of knowledge and skill and, perhaps, in being able to perform these functions better. This cannot plausibly be described as a difference of quantities of factors or sub-factors. Rather, it is the product of a past when flax was grown in these areas and illustrates the inability of the factor endowments theory to cope with experience and proprietary knowledge.

France's luxury goods industry illustrates yet another cause of trade overlooked by the theory, namely the concentrations of industries. French workers may have had more flair for such goods than English workers, acquired from the ability of French institutions to impart suitable training or from upbringing, but many foreigners drawn to France to work in the industry were as successful as the French. What drew them to France was the pre-eminence of France, more particularly of Paris. In much the same way British steel makers congregated around Sheffield, American motor car makers around Detroit, computer firms around Silicon Valley, the American film industry around Los Angeles, international financial services in London, jewellers in Pforzheim (to take an example of Ohlin's) and printers in the sixteenth century in Venice, to name a few of many examples. Concentrations of economic activities of this type are a commonplace of economic history and are determined by the advantages they afford individual enterprises. Among the reasons why an industry came to be concentrated at one site rather than another are climate, as for instance with America's film industry and Silicon Valley, and lower transport costs for raw materials, e.g. Sheffield. But often it was not a natural advantage so much as a historical outcome, as with Pforzheim and Venice or with London's financial services and the concentration of German motor car manufacturing around Wolfsburg and Stuttgart. In some cases the place in a country where an industry was concentrated had no bearing on exports. It is unlikely that the concentration of motor car manufacture around Detroit or Stuttgart made any difference to

the countries' exports, but in other cases, like the French luxury goods industry, London's financial services and the linen industry of Northern Ireland and southern Scotland, where they were concentrated did matter.

The Leontief Paradox

In 1953 Leontief compared the direct and indirect capital and labour requirements of the US's export and import competing sectors using the input-output table he had put together.⁴⁸ Although the exercise is usually depicted as a test of the two factor, two good, two country model, it can also be considered a test of the two factor version of the general form of the factor endowments theory that takes advantage of the special circumstance that the US was generally believed to have more capital per head than any other country. The results of Leontief's computations were the opposite of what was expected: the capital per head needed for a marginal increase in the output of exports was less than that needed for a marginal increase in the output of the import competing sectors.

Several ways of explaining Leontief's results while keeping the theory have been proposed. One type of explanation has been that the data are misleading and must be altered. It can take several forms. Leontief, himself, asserted that the figures for workers had to be changed because an American worker was equivalent to more than one foreign worker. From comparisons of wages, Leontief concluded that he was equivalent to three, which, as Chipman pointed out⁴⁹, meant he assumed that the prices of factors, measured in "efficiency units", were equalised across countries with appropriate assumptions about production functions and efficiency of capital.

Another such alteration of the data follows from the assertion that the superiority of the American worker is the effect of the capital embodied in him in the form of education, training, health care and so on, whereas Leontief only takes physical capital into account. If the export sectors use more highly trained labour than the rest of the economy, they may use more capital per worker without it showing in Leontief's calculations. Several economists have argued that this is shown by the higher wages in the export sectors as compared to the rest of the economy. But this does not ensure that, were the calculations to be revised accordingly, the outcome would be that of the theory. The same kind of disaggregation, quite in the spirit of Ohlin, must also be carried out for the rest of the world. Then it is to be expected that some of the exports from there that were the cause of the higher capital content in Leontief's original calculations would also require more skills to produce than other goods. Moreover, countries vary in their

⁴⁸ Leontief, "Domestic Production and Foreign Trade."

⁴⁹ Chipman, "A Survey of the Theory of International Trade: Part 3.," 56.

endowments of skills, so that the disaggregation by sector needs to be matched by disaggregation by country, or at least groups of countries – those exporting goods that need more skills to make and the others.

The only alternative, in the same theoretical scheme, to altering the data is factor intensity reversal, meaning, in this case, that one good that is produced with more capital per head than another at one set of factor prices is produced with less at another set. It is unavoidable with two factors if countries are assumed to have the same production functions but not the same factor prices. If reversal occurs once between the factors prices of the US and those of the rest of the world, the labour intensive good of the US is the capital intensive good of the rest of the world and vice versa. Then, the US can be exporting its labour intensive good and importing its capital intensive good, which is what the rest of the world will also be doing. The alternative is possible too, the US exporting its capital intensive good and importing its labour intensive good. Which alternative occurs depends on individual preferences as well as on factor endowments.

If factor intensity reversal is accepted as the explanation of Leontief's result, it must be accepted as normal to the extent of outweighing non-reversal. The US has for a long time been the biggest importer in the world and if its import competing and export sectors show reversal in the aggregate, reversals between pairs of goods must be common. Then Ohlin's notion of the relative abundance of factors is of no use.

Any other explanation of Leontief's "paradox" compatible with the factor endowments theory must be outside his scheme of two factors, free trade and common production functions. Assuming more than two factors removes the paradox by making relative abundance and intensity of use of factors ambiguous. Then a country may have a relative abundance of capital to labour and yet export goods that are labour intensive relative to capital. Leamer gives a numerical example with three factors and three goods.⁵⁰

Some economists have, nonetheless, drawn the conclusion that the modifications Leontief made to his calculations by excluding several natural resource intensive goods as "non-competitive imports" are evidence for the theory because they gave the result originally expected, that US exports were relatively capital intensive. The conclusion is unwarranted; inferences about the pattern of trade cannot be made without further justification simply by eliminating some goods or factors. Perhaps redoing the calculations taking explicit account of natural resources from the start, i.e. a different input-output table, would have yielded results favourable to the theory, but the practical difficulties were too great.

⁵⁰ Chipman, 56.

Leamer⁵¹ argues that the results from Leontief's original data agree with the factor endowments theory if the calculations are done as he proposes. He assumes the conditions for the Vanek theorem with more than two factors, which allows him to calculate from Leontief's figures and estimates of the share of the US in world income that the US was a net exporter of capital services, as theory required. To deal with one empirical result, he makes two assumptions that have no empirical justification, namely that the proportions in which goods are consumed are independent of income and that factor prices are equalised. If the paradox is resolved by that, it has changed; as Brecher and Choudhri pointed out, Leamer finds that the US was also a net exporter of labour services and, by the Vanek theorem, must have had a greater share of the world's labour than of its income, which means that the US's income or expenditure per head was below the world's average.

Deviations from free trade can, in principle, alter the pattern of trade so much as to result in trade inconsistent with factor endowments. As Chipman pointed out, the goods Leontief added to the list of "non-competitive imports" may have required large amounts of capital per head, but the natural resources were more productive abroad than in the US. The US production of such goods and many other agricultural and mineral products depended on protection against import competition and he concluded that what Leontief's results showed was that trade was so far from free as to cause more misallocation of resources than had been thought.⁵² This explanation has no obvious logical defect, but can only be substantiated by comparison with what would have occurred if there had not been protection, which cannot be known directly but would be a matter of conjecture.

Factor price equalisation

Another, less restricted form of the factor endowments theory, from which inferences about observables can be drawn, is that in which the number of goods is at least as great as the number of factors, in which case all countries have the same factor prices if the prices of goods are the same for them all and their factor endowments are not too different. If the number of equations determining the prices of goods in terms of the prices of factors is greater than the number of factors, it can be assumed that the prices of goods are such as to give a consistent set of equations. This is the first testable inference to have been drawn from the theory, for Heckscher had argued in 1919 that it would result from free trade,⁵³ as opposed to Ohlin, who asserted that trade would only diminish the differences in factor prices, but not lead to equality.

⁵¹ Leamer, *Sources of International Comparative Advantage.*, 51–54.

⁵² Chipman, "A Survey of the Theory of International Trade: Part 3.," 57.

⁵³ The translation of Heckscher's paper is in Ellis and Neal, *Readings in the Theory of International Trade.*

Samuelson first proved it in 1948 for the case of two goods and two factors,⁵⁴ though others had earlier reached the same or a similar conclusion.

The alternative of assuming that factors outnumber goods has not been popular. One reason is that it is hard to conceive that factors can outnumber goods in all their variety. A second is that recent theories that purport to explain the trade between countries in similar goods rely on factor price equalisation. For these and, perhaps, other reasons economists, when they discuss incomes in the context of international trade theory, have preferred to assume that the number of goods is not less than the number of factors.

Since wages, at least, seem to vary by factors of thirty or more around the world, the questions are, how can the factor endowments theory be reconciled with such differences if factors do not outnumber goods, and, beyond that, can the theory explain the differences? In the abstract there are three possibilities. One is that the price equations have more than one solution. A second is that factor prices may be equal, but workers are paid more than the factor price of labour. Third, countries may specialise, meaning that they do not produce all goods. Finally, trade barriers prevent trade from equalising factor prices.

That there should be more than one set of factor prices to a given set of prices of goods can be taken as normal, for the conditions for global uniqueness of the solutions to the price equations are stringent. But non-uniqueness, by itself, is not an explanation of wage differences; if countries differ in their factor prices for the same prices of goods, the differences still have to be explained. Furthermore, two countries that have the same factor prices at one time will always have the same factor prices, because the various sets of factor prices corresponding to the same goods prices are locally unique and countries cannot jump from one set to another, whereas, in reality, countries that at one time had the same wage rates and, therefore, presumably the same factor prices, have diverged later without making jumps. According to the theory, economic policy can have had no influence, the relation between factor and goods prices being the same for all countries.

If, however, workers are paid more than just the factor price of labour, factor price equalisation may be consistent with the differences in wage rates. Becker has asserted that the pay of workers consists of the factor price and the yield on capital invested in them, and that the latter accounts for the differences between countries. 'The term x represents the earnings of a person that are unrelated to human capital invested in him, and are presumably, therefore, largely independent of his current choices. Particularly in developed economies but perhaps in most, there is

⁵⁴ Samuelson, "International Trade and the Equalisation of Factor Prices."

sufficient investment in education, training, informal learning, health and just plain child rearing that the earnings unrelated to investment in human capital are a small part of the total. Indeed, in the developmental approaches to child rearing, all the earnings of a person are ultimately attributed to different kinds of investment made in him. Consequently, there is a considerable justification for the assumption that x is small and can be neglected, an assumption we make in this paper.'

So, when the unskilled worker in a developed country is paid more than the unskilled worker of the same literacy level in a developing country, ten to fifty times as much in real or money terms, the difference must be attributed to the investment in child rearing, that is to say the capital invested in the parents. The superiority of the one worker to the other seems to be hereditary, though not genetic.

Becker's 'considerable justification' is mere assertion, one that the many people who try to smuggle themselves into the US or Europe from Latin America, Asia and Africa without having inherited much investment in themselves seem not to believe. Nor can it apply to Ohlin's Italian, who functions like an average American in the US. Orthodox economists continue to assume that Becker's x cannot be neglected.

Factor prices are not equalised if the number of goods countries produce in common is less than the number of factors, i.e. countries specialise. According to the theory, countries specialise when they have too little of some factors to produce all goods at the prices set by trade. Thus, if there are two factors, labour and capital, and two goods, a country with little capital per head may just produce the labour intensive good whilst other country produces only the other good or both. Because the former country has little capital per head, its wage rate is lower than that of the others.

But this cannot be an explanation when, in reality, countries produce the same types of goods and yet have different wage rates. Wage rates in China and India are fractions of those of Eastern and Southern Europe, Korea and Taiwan, where wages are well below those of Northern Europe, Canada, Japan, the US and Singapore. In addition, comparisons can be made of the past; in the first three decades after the Second World War Wage rates in the US and Canada were higher than in all of Europe.

Free trade, protection and unequal exchange

Finally, wage differences can be ascribed to deviations from free trade, though the point at issue is not the extent to which these deviations are actually the causes, but whether orthodox neo-classical theory can explain both the deviations and how they result in the differences in earnings.

Hence, discussion of deviations from free trade is needed both because factor prices are not equalised and because of the Leontief paradox.

Until after the Second World War trade among the countries that are now developed had been free only during brief periods before the First World War and, almost without exception, the countries that practised free trade at some times had depended on protection against foreign competition at others, though this is inconsistent with the factor endowments theory, according to which a country that cannot affect international prices does best if its trade is free. Deviations from free trade are and always have been sufficient to require explanation. As Rodrik puts it, 'Perhaps no other area of economics displays such a gap between what policy-makers practice and what economists preach as does international trade. The superiority of free trade is one of the profession's most cherished beliefs, yet international trade is rarely free.'⁵⁵

Deviations are not random; they have systematic features, which is evidence that free trade regularly creates the same problems. Of these features, the main one is that the deviations have always most commonly been protection against imports. International agreements now restrict the use of tariffs and quantitative restrictions, which used to be the methods most often used, but other methods have taken their place that can be applied specifically to particular imports while being harder to contest legally, notably anti-dumping regulations, voluntary export restraints (VERs) and bilateral agreements.⁵⁶ At the least, this is evidence that the desire or need for protection against imports is persistent. Subsidies and other stimuli for exports have also been common and take many forms, though orthodox theory does not, *prima facie*, show why a country should pay to reduce the consumption of its own citizens and increase that of others.

The argument for free trade is that it allows everyone in the country to be at least as well off as with any deviation from it. For, using the Haberler representation, the country can obtain through trade any combination of goods represented by a point on the plane orthogonal to the price vector and going through the point representing the country's production. Since the country's production possibilities make a convex set, the point on the frontier of the set at which the tangent plane is orthogonal to the price vector allows the country to obtain at least as much of all goods as it can obtain by producing at any other point. If the country is assumed to have a welfare function that it maximises, free trade allows it to reach the point on the tangent plane that yields the highest possible value of the welfare function.

⁵⁵ Rodrik, *What Does the Political Economy Literature on Trade Policy (Not) Tell Us That We Ought to Know?*, 1.

⁵⁶ Haque, "The Rise of Bilateralism in Trade."

This is the common, textbook argument, of which Meade says, 'As a formal proof of the case for free trade there is really nothing to be added ...'⁵⁷.

Hence three questions arise: firstly, what explanations can orthodox theory provide for deviations from free trade, secondly, what effect do the deviations have on other countries and, thirdly, how can the optimality of free trade be empirically tested?

Deviations from free trade take many forms, so for brevity, the discussion here is confined to tariffs on imports, the form that has been most considered in the theory, both because of its simplicity and because of the extent of its use. One explanation that the factor endowments theory provides for tariffs is that, when countries produce the same goods and world markets are sufficiently competitive, each country can use them to raise its domestic prices of imports and thus increase the real income of the factors it uses relatively intensively in the domestic production of those goods. In the case of two goods and two factors, this is the assertion of the Stolper-Samuelson theorem. Even if the theorem cannot be generalised to many goods and factors without adding conditions on the production functions, it can be taken that protection against imports is to the advantage of at least one factor, which makes it a matter of income distribution.

Prima facie, any explanation that relies on the motive of income redistribution relies on the particular circumstances of the country, for it must include an explanation as to why in a competitive economy the greater income allowed by free trade cannot be reached. Various hypothetical mechanisms have been devised by which interest groups, who may have welfare functions to maximise, influence decisions on trade policy by their votes or by influencing politicians, who may also be maximising their own welfare. Such mechanisms presuppose that the losing groups accept their losses, which are at least as great as the gains of the others, because they are unaware or lack the power to defend their interests. They are either not behaving rationally or are constrained by circumstances, though all the circumstances that have been offered in explanation seem peculiar to some high wage countries at some times and not at other times and not true of other countries, whether high or low wage. Rodrik has surveyed these mechanisms and agrees with the *prima facie* view: "The results discussed above tend to be too narrow and specific to account for what is essentially a universal phenomenon: the preference of political systems to use trade intervention to generate or sustain redistribution outcomes."⁵⁸ Little can be added to Rodrik's survey if his motive for trade intervention is accepted.

⁵⁷ Meade, *The Theory of International Economic Policy*. Vol. 2, *Trade and Welfare*., 142.

⁵⁸ Rodrik, *What Does the Political Economy Literature on Trade Policy (Not) Tell Us That We Ought to Know?*, 25.

Better explanations from the same premises may yet be found, but, if a satisfactory one did exist, economists would not be speculating as to what it could be. Instead, they would be employed by interest groups and politicians to calculate the effects of different measures on income and its distribution and the optima of the welfare functions. They would have first hand knowledge and would be able to reveal it, unless the interest groups were able to do these calculations themselves and combined to keep the knowledge from the economists.

Various explanations for deviations from free trade can be given from other premises not compatible with the Haberler representation. For example, increasing returns to scale to the point that production possibilities sets are not convex, or dependence of supplies of factors on their prices, which alters the relation between the marginal transformation rates of goods and their prices. They are theoretical possibilities and planners or groups of firms or workers have at times acted according to what they thought would be of advantage to them under such circumstances, but they cannot be considered here in a discussion of orthodox theory.

A modification of the assumptions while staying within the framework of orthodox theory is that international markets are not so competitive that individual countries cannot change their terms of trade. This, too, can be modified to allow a group of countries to coordinate their protection against another group of countries. In addition, the framework of orthodox theory can be altered in two ways. One is to allow for unemployment. If the exports of some economies grow quickly enough, the competing production in other countries may be displaced and, if other activities in the latter countries do not expand fast enough, unemployment results. Temporary trade barriers to prevent such "market disruptions" were explicitly allowed in the General Agreement on Tariffs and Trade (GATT). The other modification of the orthodox theory is to allow for governments. Governments can raise revenue from taxes on trade, or they may act irrationally by assessing wrongly how they gain from deviating from free trade or by not understanding free trade's benefits.

Few orthodox economists believe that the major high wage countries deliberately alter the terms of their trade with other high wage countries to their own advantage. Any attempt by a high wage country to influence its terms of trade like this is likely to provoke a riposte, and one of the motives for the international institutions governing international trade has been to avoid such conflicts, which can end with losses for all. Rather, high wage countries show that they will accept worse terms of trade in exchange for protection against imports, as when they use voluntary export restraints (VERs) and allow the exporters to raise their prices. Such restraints became

common in the 1980s. Several hundred VERs were “official”, in the sense of having been negotiated between countries, the best known being those governing Japan’s exports of cars to the US. Bilateral agreements are usually hard to spot and some that have been negotiated between the industries of various countries may not be official, although the authorities of the countries concerned may have helped negotiate them.

But, to cope with the imports from the low wage countries, the high wage countries have used the ordinary methods of tariffs and quotas. In the 1950s and 1960s the high wage countries resorted to the GATT provisions for protection against market disruption caused by the growth of the exports of manufactures of some low wage countries. These provisions had only been intended to give countries time to adapt and were supposed to be temporary. Adaptation to the imports from low wage countries would have been easier then than it has become since, for this was a time when the economies of Western Europe were growing fast with close to full employment, world trade was growing faster still and unemployment in the US was low. Instead, the ostensibly temporary protection became permanent; governments acceded to the demands of firms and workers directly affected by the competition from low wage countries. Their motive was to prevent industries from being displaced and is not an example of the protection referred to above with the motive of increasing the income of one factor or the other. On the contrary, neither wages nor profits in the protected industries were raised above those in other activities.

Trade barriers can alter the terms of trade if the country imposing them is big enough or if enough countries importing the same goods coordinate their protection. As was to be expected, when the low wage countries began to industrialise, they mostly produced the same kinds of goods. Hence, the high wage countries coordinated their protection. Its explicit forms included the various cotton textile arrangements, trade groupings like the European Common Market and the European Free Trade Area, and the Generalised System of Preferences. The prices the low wage countries received for their exports to the high wage countries were lowered relative to the prices of the goods they competed with produced in the high wage countries. Hence the total receipts of the factors, as given by the factor endowments theory, were lowered. If the formulation of the factor endowments theory is modified to allow for tradable inputs into the production of these exports, such as raw cotton in the production of textiles, and if the prices of these inputs are not affected by the protection, the total receipts of the factors are less than the prices of the final goods and are lowered disproportionately by the protection.

This combination of a low wage rate without a correspondingly higher profit rate is also the outcome of Emmanuel’s “unequal exchange”, though

Emmanuel argued that it occurred with free trade. His thesis was that capital is mobile enough to even out profit rates around the world, but he did not explain how the prices of high and low wage countries could be reconciled. Since he put his reasoning in a quasi-Marxist form, it was dismissed or misinterpreted by orthodox economists.⁵⁹

Yet, Emmanuel's argument is easily given a valid form. Where goods exported by low wage countries compete with goods produced in the high wage countries, the assumption of free trade is dropped. In the model of Section 2 the tariff of the high wage country lowers the rate of profit on investment in the low wage country. If the producers of the latter do not start producing their own capital goods, as has been the case in low wage countries as a rule, their rate of profit can be reduced even below the profit rate of the high wage country. Emmanuel's argument can be used with free trade for goods that are not produced in high wage countries, especially agricultural products of tropical and sub-tropical climates, like cocoa, coffee, tea, bananas, papaya and jute. Trade in these can be free and the export prices are mainly determined by the labour cost in the countries producing them and the measure of unequal exchange is the amount by which the prices would be higher if the labour was paid the corresponding wages of the high wage countries. Trade can also be free for manufactures that are not produced in the high wage countries, for instance because these countries have stopped producing them and allowed the low wage countries to take over all their production. Manufactures and agricultural products like these are low wage goods.

Among those to dispute Emmanuel's thesis was Samuelson, who devised his own model, not conforming to the factor endowments theory, in which goods are made by inputs of labour over several periods.⁶⁰ It is a Ricardian model in the sense that countries do not have access to the same techniques of production and the consequent relative differences of prices results in trade. Samuelson follows Metcalfe and Steedman in assuming extraneously given rates of profit, so that the prices of goods in each country are determined by the time pattern of labour inputs and the profit rate on them. As a result, the pattern of trade and production depends on the profit rates of the countries. It is possible, then, that countries specialise at one set of profit rates in producing and exporting goods that they would not produce with different profit rates. Since the number of workers in each country is fixed and is, therefore, the constraint on output, the total output at zero profit rates cannot be exceeded. Thus, positive profit rates can result in countries producing less than the maximum, which Samuelson shows by examples. This can happen when the countries have the same profit rates, but different

⁵⁹ Emmanuel, *L'échange inégal*.

⁶⁰ Samuelson, "Trade Pattern Reversals in Time-Phased Ricardian Systems and Intertemporal Efficiency."

production techniques, but also when countries have the same production techniques, but trade because the profit rates are different.

Two points that Samuelson was making matter here. One was that, even the patterns of production and trade that gave less output than the maximum were Pareto-optimal, because the periods of lower output in at least one country as workers were transferred in the transition to maximum output could not be avoided; an immediate loss of output of one good would be followed by a delay of one or more periods before the inputs of the transferred workers resulted in an output. His second point was that Emmanuel's thesis was that free trade resulted in "deadweight loss". Samuelson seems to have drawn this conclusion by interpreting Emmanuel's assertion that free trade was not optimal as meaning that free trade was not Pareto-optimal.⁶¹ Emmanuel did not, himself, refer to Pareto-optimality and denied that he had argued that free trade caused a deadweight loss.⁶² His statement that free trade was not optimal was a value judgement about the distribution of income between high and low wage countries, one that many would agree with, and in keeping with the tenor of his book.

Samuelson's procedure of devising a model to yield a different conclusion does not, by itself, refute the conclusion he contests. What his choice of model for this occasion indicates is that the standard factor endowments theory, as exemplified by his formulation of 1953, is unsuited to discussing Emmanuel's thesis. If there are several factors, the price of one factor in a country cannot be lower than elsewhere without the prices of other factors being correspondingly higher, unless trade is not free or the country is completely devoid of all but one factor and is specialised in the production of goods not produced in the other countries. Samuelson's depiction of production techniques as inputs of labour over a succession of periods before the final product would have been akin to having fixed capital goods but for the assumptions that the inputs of labour were only used once; durable capital goods are excluded. The labour inputs of the various periods cannot be traded either. Implicit in Emmanuel's argument was production by workers with a stock of durable capital equipment, which may or may not be tradable, and a rate of profit on the value of that stock.

⁶¹ Samuelson, "Free Trade's Intertemporal Pareto-Optimality."

⁶² Emmanuel, "A Note on 'Trade Pattern Reversals.'"

DIFFERENTIATED GOODS, BRANDS AND TECHNICAL PROGRESS

1. DIFFERENTIATED GOODS AND BRAND NAMES

Complete and incomplete explanations of trade

Explanations of the pattern of trade, i.e. which goods countries export or import and in what amounts, can be of two kinds. In the one the determinants of trade are quantifiable characteristics of countries and how they compare, whereas in the other the characteristics that determine trade are not quantifiable. Theories of comparative advantage and the models of Bensusan-Butt and Chapter 1 are of the first kind. Such theories and models are complete in the sense that they determine each country's exports, imports and production. According to explanations of the second kind a country exports a particular good because of characteristics that are not quantifiable and which do not necessarily permit of comparisons with other countries. Thus, France has long exported perfumes because of circumstances in its past, even centuries ago. This does not prevent another country with different characteristics from also being a perfume exporter because of circumstances in its past. Similarly, Venice and then Holland were once centres for printing and Hollywood remains the biggest source of films among high wage countries. In each case the original reason, commercial connections, social structure, political toleration and climate, was not quantifiable. Explanations of the second kind cannot be complete; the reason a country exports a particular good may be a characteristic of the country, but the quantification of characteristics needed for comparison between countries, as with, say, the factor endowments theory, is omitted.

An explanation of trade that is complete in the sense given here may be too much to ask for. Theories and models intended to provide such explanations necessarily depend on the differences between the quantifiable characteristics of countries, and there are two common phenomena that cannot be reconciled with that. One is that countries export and import the same goods and the other is that countries with different wage rates produce the same goods. Neither should occur, except by coincidence as exceptional cases. As examples of both phenomena, France, Germany, Japan, Mexico,

Sweden, the UK and the US import and export cars and car parts, France, Italy and the US import and export clothing and cosmetics, Brazil, Canada, a group of European countries and the US exchange civilian aircraft and most of the bigger industrial economies import and export industrial equipment and intermediate goods similar to some they export.

In the factor endowments theory, since the goods in question are produced at the same prices in the different countries, either phenomenon implies that the number of factors is greater than the number of goods, which, given how many goods are traded, is implausible. Furthermore, since the number of different traded goods has been increasing since the start of the industrial revolution, either the number of factors has been growing too, which would have to be explained, or it was already bigger in the past than the number of goods at present, which would also need explanation. Ricardo's theory and its extension to several countries and several goods are incompatible with either phenomenon since both require special assumptions to allow two countries to produce more than one good in common. Put heuristically, if trade between two countries is to be explained by comparisons of the countries' characteristics, either the number of characteristics is large or few goods are produced in both. If only a few characteristics determine the relative prices all goods would have in a country if they were all produced there, few goods would have the same relative prices in both countries, except by chance. But when more than two characteristics are compared, comparison of their relative amounts is ambiguous unless a common standard is specified and the existence of such a standard requires additional conditions. This was seen in the earlier discussion of the factor endowments theory and factor price equalisation.

Incomplete explanations with differentiated goods

An illustration of the difficulty of formulating a complete theory that allows for these phenomena is given by Helpman and Krugman⁶³, who, in Chapter 7 of their book, try to account for the first phenomenon by including product differentiation in the factor endowments theory and end with an explanation of the second kind. Differentiation here means that a good has several versions, the differences being a matter of design. Motor cars illustrate this; there can be thousands of differences between two cars, though both are the same good. Differentiation in this sense is common among manufactures and processed consumer goods, but producer goods are also differentiated according to the specific needs of the user and often made to order. Competition consists of buyers choosing, according to their preferences and incomes, from the different versions of a good according to the characteristics and prices. Producers distinguish themselves as brands and their versions by

⁶³ Helpman and Krugman, *Market Structure and Foreign Trade*.

names, which allows them to try to influence buyers through advertising, and it is possible for versions to differ only in name and appearance. Establishing a new brand or maintaining an old one normally requires advertising and similar selling costs, which can be as big a part of total cost as profit or R&D. However rational or irrational the effect of selling costs on consumers may be, the price and supply of a version of any good is determined by the producer's assessments of buyers' preferences and the prices and characteristics of competing versions. As a rough rule, the more elaborate the manufacture, the greater the variety of specifications and the more branding there is. And, as a further rough rule, elaborateness is indicated by the difference between the cost of the raw materials and the price of the good. Goods that are not differentiated are referred to here as homogeneous.

Helpman and Krugman assume that all countries have the same factor prices, which, according to Vanek's theorem, allows them to export and import the same goods. If, to make their argument more plausible, the assumption of equality of factor prices is confined to some countries only, not all, it need only be assumed that the theorem's conclusion can be extended to sets of countries, each set consisting of all the countries with the same factor prices. Then, if there are increasing returns to scale in the production of each version separately, as opposed to increasing returns for the total production of the good, the competition of producers in one set of countries makes them confine their production to versions that other producers do not make and the versions a country does make are not the same as the ones it imports. Each version of a good is the monopoly of the producer, which allows the price to be higher than it would be with perfect competition and, thus, avoids the problem that increasing returns causes at least one factor to be paid less than its marginal product. It also prevents arbitrage in which a version of a good is bought at the competitive price and sold to undercut the monopoly.

The explanation that results is of the second kind because the relevant quantifiable differences between countries are assumed away. Increasing returns have the appeal of seeming concrete, but this formulation involves no quantification and no comparisons across countries. Any other mechanism that causes firms with the same factor prices to produce several versions of a good does as well, provided there is something to prevent two firms from making the same version; if there is, the countries in which these firms operate will also be exchanging different versions of the same goods.

It is also a static, single period argument, as is all the discussion of the book, despite the assertion, 'To the extent that a static model is used as a proxy for a dynamic world, it should be viewed as a representative of the

whole time path of that world, not a snapshot at a particular point in time.⁶⁴ Yet it can neither represent a time path nor be a snapshot. For a time path the same country in the different periods would either have to be represented as different countries, each with its own set of factors, or it would have to be represented as one country in one period, but with its factors of different periods being represented as different factors in that period. In either case there would be a set of relations among factors of different periods. For instance, the factor, capital, can be increased by investment, the labour force can grow naturally and the supply of land can diminish because of erosion. In each period the amount is related to the amount in the previous period. These relations are not allowed for; the number and quantities of factors are fixed. In particular, Helpman and Krugman have no saving or investment, another reason their model cannot represent a point in time. Most of the book assumes that all countries have the same factor prices, which implies either that factor prices do not change or that they change in the same way in all countries.

With these limitations it is not to be expected that the differentiation mechanism the theory derives will be what is usually observed. It is closer to the reverse. Other things being equal, firms, according to this theory, specialise in fewer versions because that reduces the unit cost of production, whereas what is observed is that firms try to offer more versions rather than fewer. Most often a firm trying to expand its sales of a particular good produces new versions that compete more closely with versions produced by other firms than do the versions the firm has so far been making. This would be consistent with increasing returns if it were applied to each firm's total production of a good, not of each version, but that would eliminate the mechanism of Helpman and Krugman.

Heterogeneous, durable capital goods do no better than the factor endowments theory at explaining this kind of trade when consumption goods are assumed to be homogeneous and all countries to have access to the same production techniques. With these assumptions, in any single country, when the real wage or the profit rate is given, the relative prices at which goods can be produced there are specific to that real wage or profit rate. The same goods produced in countries with different wage rates and in which wage and profit rates are uniform must have different relative prices. So, if two countries trade freely, they specialise, which is to say they produce at most one good in common, unless the wage or profit rate adjusts to allow two goods to be produced in common. Only by chance would the relative prices of more than two goods be the same in both countries.

⁶⁴ Helpman and Krugman, 39.

Moreover, it is practically impossible to give precise formulae for many quantities, like prices and output. Unless the economies are static or all grow steadily at the same rate, the pattern of production and trade changes with time and, unless economies function with perfect foresight and coordination, i.e. an intertemporal general equilibrium, such change implies either that profit or wage rates are not uniform within countries or that the law of one price does not hold. In the model of Chapter 1 the law of one price held and, when perfect foresight and coordination were assumed, the cloth producers in the high wage countries reduced their output, raised the price to compensate for the shortening of the life of their capital equipment and shifted to other activities. But, otherwise, the growth of production in one country caused profit rates not to be uniform in the other. Similarly, a relative change of the nominal wage rates of various countries, say because of a change of exchange rates, alters relative prices and profit rates. If there was equilibrium as described, it ceases and again the law of one price and the uniformity of wage and profit rates within countries become incompatible, though it is possible that neither holds. After repeated changes, especially of exchange rates, the relations of profit, wages and prices in the various countries become complicated and unpredictable. Ricardian and most factor endowments models allow precise formulae because they are confined to single periods, though there are factor endowments models that include several periods and yield precise formulae because they avoid the necessity of equating the costs of production of capital goods and the profit from them.

Precision is confined to models in which the outcome is equilibrium, meaning that expectations turn out to have been correct, and can never come about in practice. Multi-period models with heterogeneous, durable capital goods can, perhaps, be so devised as to be shown, by the use of topology, to have intertemporal general equilibria, in which case they have precise outcomes, though none seem actually to have been devised for international trade. But explicit formulae for prices and output have never been derived for intertemporal models that have several goods, people and firms, with or without international trade. Were the formulae to be derived, they would be intractably complicated. Nor do such models include explanations as to how individuals or firms are to know the formulae and use them, hence how they know how to act consistently with all the others. Single period models, like that of Helpman and Krugman, necessarily have equilibria, for there are no additional periods for expectations to turn out to be wrong or for inconsistencies in the intentions of firms and individuals to have consequences. All firms and households in these models must be supposed to act mutually consistently as though by agreement and, apart from the possibility of several distinct solutions to the equations, their behaviour is fully determined.

Any attempt to describe or explain international trade realistically, therefore, must allow that households, firms and government cannot have exact knowledge of prices and demand in the future, let alone what new goods or versions of goods may become available. Under normal circumstances they can have approximate knowledge; if there are no drastic, unforeseen changes, they can rely on experience, publicly available information and their own special knowledge to know roughly what to expect. Experience substitutes for the coordination of households, firms and government that would occur in equilibrium, when all have the exact information about the present and the future needed for decisions. The greater the unforeseen changes, the more outcomes differ from what had been expected and the more uncertainty there is about the actions of others. A change can, for example, cause unemployment and the reaction of some households may be to forego consumption just in case; they react to not knowing who will become unemployed. A change foreseen by some but not others has the same effect of creating uncertainty about how the others will react.

Hence, assertions about how economies are affected by specific events can only be made by stating how households, firms and government are assumed to assess the events and react to them. There is no determinism, assessments and behaviour can change at each stage. Ruling implausible and irrational behaviour out can reduce the range of possibilities to a few cases of interest, which does not prevent reliable statements about the future, but limits them to statements about what alternatives can follow from a given situation, and that is the most that can be hoped for or wanted if realism is the objective.

Nonetheless, there is no alternative to the assertion of Helpman and Krugman, that the explanation of why countries often import and export the same goods is product differentiation. There is no reason why countries should import and export a homogeneous good, except transport costs. Such cases have occurred. An example as early as the nineteenth century was the import into northern France of coal from Germany and the export of French coal to southern Germany because it saved transport costs. But this cannot explain a country's imports and exports of the same goods when the trade occurs over short and long distances. In the following transport costs are ignored or regarded as insignificant.

Product differentiation does not, by itself, cause this kind of trade; the same versions of goods as made abroad can be made in the country. If a country imports some versions of a good whilst its firms make other versions that they export, it must be supposed that there is an obstacle to firms of the country making and selling the imported versions at competitive prices. One possibility is that some versions are cheaper to produce where the wage is higher and the profit rate is lower and others more expensive, much as though they were

different goods. It may occur often but most of the trade between countries of the same goods is obviously not of this kind. Moreover, versions of goods change often compared to wage rates, but not their places of manufacture. Hence, there are obstacles and they must either be legal or technical.

Branding as obstacle to the law of one price

One obstacle is that the laws governing trademarks and intellectual property give exclusive rights to brands and prohibit copying versions made by others without their consent, at least for some specified time. Most manufactures are sold with brand names and a version of a good produced under one brand cannot be produced without the permission of the owners of that trademark as long as the prohibition is in force. Versions of goods for which the prohibition has ended, as with the expiration of patent protection, or which existed before these laws applied may be produced by all. Goods like steel, tea and paper have several varieties that have existed for long enough that any firm may produce them, but a firm producing a new variety normally tries to prevent imitation by the use of these laws. Brand names do not necessarily imply differentiation; a homogeneous good may be sold by different producers under various brand names. But, when differentiation is accompanied by the application of the trademarks and intellectual property laws, it is accompanied by branding.

Apart from its necessity for producers, branding serves a purpose for buyers. Manufactured goods can vary because of design, including the process of manufacture, or, given the design, because of the quality standards in making them, and for buyers the brand name is an indication, if not a guarantee, of both. Durable manufactures often need maintenance and their owners may want assurance of qualified repair services and a supply of reliable spare parts, all of which the manufacturer may need to provide directly or indirectly, at least as long as the guarantee lasts, though the owner may pay for the services for longer. The more these things matter the greater the importance of the brand name. Homogeneous goods for which these things are unimportant can sometimes be differentiated by branding, which is to say that brand names influence buyers enough to affect prices, but the availability of the same goods with other brand names or without any limits the effects. Brand names do not necessarily imply a specific producer and a producer can produce for several competing brands. But it is through the brand and the seller that the law imposes quality requirements, including partial guarantees where people's safety is concerned, as with food, medicine, buildings and vehicles. This cursory discussion of branding is all that is needed here. For brevity the term "branding" will be used when discussing differentiated goods of which versions produced by one firm may

not be produced by others without agreement of the owners of the relevant trademarks or patents.

Branding as described here prevents price arbitrage between countries. Arbitrage depends on the buyer being indifferent as to the supplier and the supplier as to the buyer, and branding allows neither. Arbitrage is also excluded from intra-firm trade, deliveries of unfinished goods produced by a firm's subsidiary in one country to other parts of the same firm in other countries, which now accounts for a large part of all international trade.

Consequently, when the prices of the same version of a branded, differentiated good differ from country to country, the differences can only be reduced or eliminated by the suppliers. Since the prices of different versions of the good in the same country are also related by competition, for instance the prices of versions that are more alike being more closely related, they can also be inconsistent between countries in this sense. That is to say, the relative prices of different versions of a good need not be the same in different countries. Such inconsistencies also can only be reduced or eliminated by the suppliers.

This cannot happen under the factor endowments theory, in which flexible prices, malleable capital and mobile factors, allow a firm to divert sales of its goods from where it has been selling to where it expects higher prices. Production adapts instantly, factors shift from producing goods whose prices have fallen relatively to producing more of goods whose prices have risen, factor prices remain uniform in each country and price inconsistencies between countries do not last.

It can happen when capital goods are heterogeneous, durable and not freely transferable between uses. To begin with the case of a country that meets a large part of its demand for a branded, differentiated good from its own production, if the pattern of trade is not changing, the prices of the several versions its firms sell there are determined by the wage and profit rates of the country. Foreign firms, therefore, price the versions they sell in that country according to the versions produced there, with which they compete. In other words, they price to market.

If the country's prices of the good were to rise relatively to prices in other countries, foreign firms would have an incentive to increase their sales in that country. In the case that the price rise accompanies an increase of demand, both domestic and foreign firms can increase their sales if they can increase their supplies there and there is no *a priori* need for the relations between the prices of the various versions to change. In the case that the price increase does not accompany an increase of demand, for example if it is the result of appreciation of the country's currency, foreign firms that can

increase their supply have an incentive to increase their sales and reduce the sales of domestic producers by pricing their versions of the good slightly below the prices that would leave the market unchanged. Domestic producers can try to match the lower prices until a point is reached at which they cannot go lower, because they cannot cover their wage and material input costs, or their foreign competitors stop reducing their prices. In either case the pattern of trade is changing and foreign firms price to market; they set their prices according to the prices or costs in that country.

Unless they can increase output with existing production capacity, the only ways the foreign firms can increase their exports to the country where the prices have risen relatively are diversion of sales from other markets or increases of capacity. Either has a cost and a firm's decision will depend on its judgment of the likelihood of the change lasting long enough to justify the cost. For instance, a price rise because of appreciation of the currency will be reversed if the old exchange rates are restored. Conversely, it is possible that the appreciation results in the elimination of positive profit margins from the production of some goods in that country, but local firms continue production and accept losses for a while because they expect the appreciation to be short lived.

Branding gives rise to costs because establishing a brand entails outlays on marketing, on creating sales organisations, such as wholesaler agreements, retail agreements or authorised distribution chains, and outlays on service arrangements for repairs, spare parts and technical advice. In addition, these things take time to arrange. When a brand is important it is significant to the buyer and establishing it is specific to the country. So, it is to be expected that establishing the brand in the market, i.e. marketing, has a cost and branding is more likely to be unavoidable, especially for consumption goods, the fewer established brands of that good there are. Some goods may have few brands, in which case, if it is not the technical aspect of production that limits the number, it must be the expense of establishing the brand that does. Other goods may have many brands, though, as with garments, it is possible that a few are more significant to buyers than the many others. The final prices of a good, therefore, include associated marketing and distribution costs, akin to capital goods in being sunk costs and not transferable to other uses, and much, possibly all, being untradable goods and services.

Hence, there are costs to gaining new markets or increasing sales in existing ones. There are also costs to regaining markets from which sales have been diverted. A firm that diverts sales incurs costs in trying to sell more in the market where prices rose relatively and, when the gain is transitory, in trying to regain markets where it reduced sales.

Diverting sales is the quickest way of profiting from higher prices, but firms outside the country where the prices of a particular branded, differentiated good have risen relatively can be expected to want to increase their output of that good if they think the prices are likely to stay higher. They can increase output by adding to their domestic production capacity, by creating capacity in the country in which they wish to sell the additional supply, create capacity in some third country or arrange with other firms in their own or other countries to do some or all of the manufacturing, i.e. subcontract. These alternatives can also be combined.

Any increase of a firm's production capacity requires investment and has a cost. If the firm's exports of the good to the country where prices have risen displace production in that country, it is unlikely that an investment there will be justified since the cost advantage that allows the local producers to be displaced is lost. If, for example, the currency of the country has appreciated, the cost of production there will have risen for all producers. It may, nevertheless, be prudent to set up production capacity there if the market is sufficiently profitable and there is reason to believe that its authorities may use trade barriers to protect domestic production. Investment in third countries or subcontracting may be more profitable if wages in these countries are lower, or may be preferred because the exporting firm's own country is at the limits of its production capabilities and investment there is hindered by shortages of workers.

Empirical: pricing to market

Pricing to market as described here means that the prices of imports are matched to prices or costs in the country, which has so far meant that the country produces the good in question and that these prices are set by the country's costs of production. Nothing need be said here about pricing to market of a good in a country where it is not produced, such as motor cars in Denmark or Switzerland. It implies that the prices of the same versions of a branded, differentiated good can differ from country to country and, therefore, that the prices of branded, differentiated goods in one country can be inconsistent with the prices in another country. Then, when exchange rates change the price indices of countries move to some extent relative to each other with their currencies, the extent being greater the greater the weight of branded, differentiated good and untradable goods in the indices, at least in the short run. There has been much empirical work on such relative movements of the price indices of various countries and some of it is discussed in what follows.

As mentioned above, firms in some countries producing a particular branded, differentiated good can increase their exports to another country so as to displace entirely competing local producers while pricing to market all

along. This is the process described in Chapter 1. Their competitors in the importing country may be displaced slowly over decades or quickly in a few years. An example of quick displacement was that of German camera makers by Japanese in the late 1950s and early 1960s. The transition from the worldwide dominance of the former to that of the latter only took a few years and has been permanent. Judging by the speed at which it happened, this could even have been a case of not pricing to market, with Japanese firms setting their prices lower, in which case they may have foregone profits. In contrast, the displacement of US car production by Japanese imports was slower and was eventually stopped by protection in the 1980s. Since then, apart from the financial crisis that began in 2007, which was not caused by imports, the motor car market of the US has been stable, the shares of the various brands of cars not changing by large amounts in a short time.

An illustration of pricing to market when the market is stable is the case discussed by Krugman in 1987⁶⁵ of German cars imported into the US as the dollar rose and fell during the 1980s. From 1980, when it averaged DM1.82 (Euros 0.93), the dollar rose year by year to reach DM 3.31 (Euros 1.69) in March 1985, after which it declined every year but one to reach DM 1.62 (Euros 0.82) in 1990. But the dollar prices of German cars in the US did not follow the German mark; their relation to US car prices changed little, some even went up as the dollar rose. For a time the price of a German car in Germany was about half the price of the same model meeting US specifications in the US. Some makes of cars sold models meeting US specifications in Germany to residents of the US, for whom it cost less to fly to Europe, buy the car, use it and then have it shipped than to buy the car in the US. Arbitrage would have had to take either of two forms. One would have been to buy many cars of a particular model in Germany and ship them to the US, which would have run the risk of failure because customers would not have had their choice of specifications (engine size, automatic or manual, colour, sun roof, seat covers, etc.) and would have no maker's guarantee. The other would have been to take orders in the US and then buy the cars in Germany, that is to say, displace the authorised dealers in the US, something the company producing the car had no incentive to allow, both because its cars were selling as fast as they could be produced and because it had no assurance that the arbitrageur would take the measures necessary for keeping the brand's reputation.

Since then it has become a commonplace that the prices of cars vary from country to country. Goldberg and Verboven, for instance, compared the prices before tax of several models of cars in various European countries over the years 1993-2003 using European Commission (EC) data and found that, 'Price

⁶⁵ Krugman, "Pricing to Market When the Exchange Rate Changes."

differentials are easily 20% or larger through the entire period.⁶⁶ The units for the comparisons were the prices in the Netherlands. Such differences even occurred between neighbouring countries using the same currency.

Motor cars are merely a conspicuous case of such price disparities, which, according to the argument made here, are to be expected for most differentiated goods. But is it the factory gate price that varies according to the destination of the export, only the final price before tax or a combination of the two? From comparisons of export prices, which can be assumed to be close to factory gate prices, it seems to be the last, which is to say that the export price and the difference between export price and price before tax to final user are both variable.

That export prices change with the exchange rate relative to final user prices in the importing country was shown for a number of goods by Isard in 1977 in what was perhaps the first systematic, empirical attempt to check directly whether the LOP held or not.⁶⁷ He calculated the ratios of indices of German export prices to indices of US wholesale prices for nine categories of goods for eight three month periods from 1968 to 1975. Exchange rates were fixed until 1973, though the German mark was revalued several times after 1968, and were then left to float, whereupon the dollar price of the German mark rose during 1973 and fluctuated around the higher rate after that. All nine of the price ratios roughly followed the exchange rate; they rose with it until 1973 and all ended higher than at the start. But five ratios fell after 1973 so as to end with increases well below that of the exchange rate, the highest being that of "glass products" at just above half the increase of the exchange rate. The other four were "industrial chemicals", "agricultural chemicals", "plastic materials" and "paper products". "Metal working machinery", "electrical industrial equipment", "home electronic equipment" and "apparel" all ended higher than the exchange rate. Isard calculated similar price ratios for six categories of machinery for the month of June over 1970–75, all of which rose at least as much as the exchange rate until 1973, but ended the period with increases that, though less than that of the exchange rate, were 70 per cent of it or more, the one exception being "agricultural tilling machinery", of which the increase was less than half.

Knetter obtained results similar to Isard's.⁶⁸ He estimated statistically for each of four countries, Germany, Japan, the UK and the US, a relation between changes in the prices of exports and changes in the exchange rates between the currencies of the exporting and importing countries (deflated by the price index of the latter) over the 13–15 years up to 1987. He added the restriction

⁶⁶ Goldberg and Verboven, "The Evolution of Price Dispersion in the European Car Market."

⁶⁷ Isard, "How Far Can We Push the 'Law of One Price'?"

⁶⁸ Knetter, "International Comparisons of Pricing-to-Market Behavior."

that in each country the relation be the same for all countries to which the exports went. It follows that if a country exports the same good to two countries and its exchange rates with them change in different directions, at least one of the export prices cannot be the price in the home market. Knetter's results were that the export prices changed so as to offset partially the change in the exchange rate in 14 out of 19 cases in Germany, 8 out of 14 in Japan, 5 out of 9 in the UK and 5 out of 11 in the US. 'However, for those industries in which exact matches are possible across countries, behavior is remarkably similar. Consequently, industry effects appear to be more important than source-country effects in explaining the dispersion ...'.⁶⁹

This rules out the possibility that producers price exports and goods for the home market in the same way, i.e. that final user prices in Germany normally varied with prices in the US. The observations referred to earlier show that they did not do so in the case of cars, and Knetter's results show more systematically, though indirectly, that this is common among manufactures.

Isard's and Knetter's results show that the difference between the export price and the price to the final user in the producing country is variable. The same can, then, be supposed for the difference between the import price, i.e. landed cost including import duties, and the final user price in the importing country. Import and export prices (f.o.b., f.a.s.) are convenient to use because they can be obtained from the countries' trade statistics as prices or unit values and are to some extent interchangeable since shipping costs can be assumed not to vary with the price of the good (i.e. variations in insurance can be ignored). Hence, changes in the export price are the same as changes in the price at import, though if a country both imports and exports the same good its import and export prices can change in different ways. But between the import price and the final user price are taxes and distribution costs, which differ between countries. They differ within a country, too, according to whether a good is locally made or imported. Hence, changes in export/import prices do not necessarily change final user prices by the same amounts and the rises and falls in the ratios of export prices to wholesale prices in the importing country that Isard calculated could have been offset by reductions and increases of distribution costs so as not to change final user prices. For example, if the currency of the importing country appreciates relative to that of the exporting country, the final user price of the good in the importing country need not change; at one extreme the export price can stay constant in terms of the currency of the importing country and, at the other, it can stay constant in terms of the currency of the producing country. In the first case the importers' profit margins rise by the full amount, allowing

⁶⁹ Knetter, 474.

for taxes, and in the second the gain goes entirely to the producer. Producer and importer may have profit sharing agreements, or the former may own the latter, wholly or partly, and import/export prices may be determined both by bargaining between importers and producers and by comparisons of direct and indirect taxes in the two countries.

Consequently, if short term variations of exchange rates raise cause gains and losses shared between producers and their authorised agents or distributors in the importing country, statistical tests of the hypothesis of pricing to market that use export prices may confirm that the prices of exports to the US vary with the price of the dollar, but will underestimate the extent of pricing to market. For instance, Knetter's results imply that expensive German cars were not priced to market in the US during the 1980s, though the common observation that they were was what prompted Krugman to discuss the matter. In contrast, the reason why Knetter's results imply that American cigarettes, British books and whiskey and Japanese fish hooks were not priced to market either may have been that these are goods for which price arbitrage, diversions of sales and increases of supply were easy.

In the foregoing the term pricing to market can refer to final user and to export prices. When considering the final buyer's choice between a domestically produced version of a good and an import, the price that matters is the final price, which is set by the distributors, the sales or value added tax being given. This is how Krugman and Goldberg and Verboven use the term. When considering the producer, the variation of prices according to the country to which the good goes makes the term applicable to factory gate or export price. Knetter and Isard refer to export prices, as do Marston and Yang, both of whom have found statistical tests to give evidence for pricing to market in that sense. The two definitions are mutually compatible, though neither implies the other since there are taxes and distribution costs in between. As the case of cars shows, the second definition does not imply the first and, it is possible that an appreciation of the currency of the importing country relative to that of the exporting country results in the export price of a good rising by, say, half the amount of the appreciation and the final user price falling while the difference between the two increases.

All this raises the question, does the normal functioning of economies remove such price inconsistencies sooner or later? The starting point of this discussion was the theoretical point that, if countries with different nominal wage and profit rates produce a variety of goods in common and capital goods are heterogeneous, durable and not transferable between uses, the relative prices of the goods will not be mutually consistent. There has been much empirical work showing that this is what happens in practice and some of that work has been described here. Nevertheless, it can be supposed that,

if exchange rates, nominal wages and the versions of goods available were never to change, countries would gradually come to specialise so that prices of tradable goods become uniform and mutually consistent across countries and profit rates uniform within countries. It might happen faster if firms were to invest in production both inside and outside their own countries.

But the versions of goods on offer change all the time; part of technical progress, probably the greater part, consists of the provision of new versions of existing goods and of new goods and the discarding of old versions and goods. This is true of capital, intermediate and consumer goods. Nothing can be said as yet about where any new version of an existing good or a new good will be produced and, so, no statement can be made as to whether or not technical progress will reduce or increase the mutual inconsistencies of prices. It is possible that the existence of technical progress prevents or hinders economies from eliminating price inconsistencies and the more so the faster the technical progress.

This is in the abstract, but there have been other obvious reasons why economies are in flux. Among them the disparities of countries and the remoteness of any stable pattern of specialisation; the pattern of trade would continue to change for a long time, even if there were no population growth, environmental problems or limits on natural resources. But the price differences of the high wage countries seem to diminish comparatively quickly. Some of the statistical studies of changes of price indices and exchange rates, i.e. of real exchange rates, using more up to date methods conclude that the price differences diminish by half in three to five years, slow when compared to what has been assumed by orthodox theory, but, perhaps, fast when the repercussions of changes in the world's pattern of production and trade and the advent of new goods and versions of goods are taken into account.

Under the right circumstances price arbitrage can occur even for motor cars and the result is greater uniformity of prices. In the European Union (EU), which is meant to be a unified market for goods, price differences as described by Goldberg and Verboven are considered to be inconsistent with its principles. Similar price differences have and probably still do occur for other goods, but the motor car is the biggest single expenditure on a tradable good of most households and attracts attention. At first the European Commission (EC) exempted the industry from some of the competition regulations, presumably because it was aware that the price differences between countries were not merely a sign of monopolies. There were rules limiting the price differences of any single model of a car, but they were not enforced systematically. In time the Commission began to foster competition in the sale of cars in different countries by making arbitrage easier, notably

by making it possible for authorised dealers to sell the same models in different parts of the EU. Some of the restrictions that car makers used to impose on repairs and the sales of parts were also prohibited.⁷⁰

It seems that as a result car price differences fell compared to what Goldberg and Verboven described. Arbitrage was probably the main reason and may partly have worked through expansion of dealerships, but, apart from that was the willingness of individuals to travel. Germans at one stage travelled to Italy to buy German cars because of the price difference and the EC in 1998 stopped the German car maker from preventing dealers in Italy from selling to buyers not resident there. Perhaps because price differences are smaller and perhaps because it is aware that there can be price differences between countries whilst competition is great enough for some car makers to have difficulty surviving, the EC has stated that it believes the car industry to be highly competitive.⁷¹

These results were possible because there was a single authority regulating competition for the whole EU and the principle that the union was a unified market was accepted by all. One aspect was that the technical specifications were the same for all the countries, leaving aside those depending on which side of the road cars drove. This seems to have sufficed for the changes of regulation, which were spread over a few years, to have had their intended effect. Arbitrage in the ordinary sense of somebody buying cars wholesale in one country and selling them in another seems not to have occurred. It would have required buying from the dealers and was unlikely to be profitable.

In contrast, there were no such changes of rules for trade between Europe and the US. During the period in the mid-1980s when the US dollar was exceptionally high relative to the currencies of Europe the authorised dealers of European cars in the US did not take to arbitrage. As already mentioned, some car makers in Europe sold cars that conformed to American specifications to US residents, whereas others refused to do so except through their authorised dealers in the US. There was a so-called grey market for the latter's cars, in which US residents bought models conforming to European specifications and had them converted in the US to American specifications, thus losing the maker's warranty but saving enough to make that and the costs of the journey and conversion worthwhile. Those car makers who allowed US residents to buy cars conforming to American specifications in Europe presumably sold the cars at roughly the export prices for the US, which, judging by Knetter's results, were higher than the local prices. They were taking some of the dealers' margins. In either case, this was the only

⁷⁰ Commission Regulation (EC) No. 1400/2002. 31/07/2002.

⁷¹ EC press release. IP/10/619 of 27/05/2010.

form of arbitrage and if the number of cars was too small to make a difference to the prices dealers set in the US.

2. SIMPLE GOODS, COMPLEX GOODS AND TECHNICAL PROGRESS

According to the argument of Chapter 1, when trade is free firms in low wage countries obtain the same incomes per unit of output from producing the same goods as firms in high wage countries, assuming they are equally efficient, though, the proportions of income going to profit are different. If there is no FDI, industries in the low wage countries grow as fast as saving rates allow and, if there is FDI, much, if not all, of the saving of the high wage countries is invested in setting up production capacity in the low wage countries, where the profit rates are higher.

In reality what were the low wage countries in the early years after World War II have differed in their economic development: Japan, Taiwan and South Korea have become high wage, technically advanced economies; some, like China and India, are still low wage countries, but becoming technically advanced with rising incomes; several, like Malaysia, Thailand, Tunisia and Turkey, are low wage countries that had rising income per head but are not at technical levels comparable to the ones already mentioned; many others are like Pakistan, they have remained low wage countries with income per head rising slowly at best and technically backward. Moreover, foreign direct investment (FDI) has not followed the pattern described here: it was not until the 1970s that it began to grow in importance, then, apart from the extraction of raw materials, it was concentrated in a few economies and, when they began to industrialise rapidly, the first three economies mentioned as a rule kept it out.

What follows is an attempt to use the argument of Chapter 1 to derive a schema from which some reasons for these differences can be deduced. It excludes the communist countries, to which the argument does not apply, and countries with negligible manufacturing industries but high wages from the production of primary products like oil. All countries are assumed to have been at the start either high wage and developed or low wage and undeveloped. Practically all countries could be categorised in this way from immediately after World War II or when they became independent until the 1970s; manufacturing in all low wage countries, except Japan, consisted, until then, of little more than producing a few consumption goods and processing of primary products, whilst other manufactures, including practically all tradable capital goods, were imported from the high wage countries.

The procedure is to find out which assumptions do not hold, beginning with the assumption of free trade. It was pointed out in Chapter 1 that, if the

high wage country decides to protect its industries against the imports of the manufactures of the low wage country, it reduces the income the low wage country obtains from producing those goods. This is what happened; after World War II all high wage countries protected their manufacturing and for several decades, as they reduced barriers to trade among themselves, they kept or increased the barriers against several of the manufactured exports of low wage countries, including textiles, garments and leather products. Trade in these goods was regulated in detail through the various textile and multi-fibre arrangements and the Generalised System of Preferences.

Stating the question

A conclusion of Chapter 1 was that, if trade barriers of high wage countries lower the prices that low wage country producers receive by enough, these producers obtain higher rates of return on investment from making their capital equipment themselves, whether for domestic use or for export to the high wage countries, than from importing the equipment and making the good. This is obviously not peculiar to the model used, but a consequence of the assumption that all countries have access to the same techniques of production, i.e. that a producer in any country can in any period produce the same amount of the same good using the same equipment and the same number of workers as a producer in any other country. It is a conclusion drawn from the low prices and does not depend on the cause being trade barriers. Yet, from the Second World War on, most low wage countries have tried to obtain foreign exchange by exporting goods also produced in the high wage countries, which imposed trade barriers, or by exporting goods that were not produced in the high wage countries and fetched lower prices than they would have if their wage and profit rates had been those of the high wage countries.

Hence, the question that has to be answered is, why have so few low wage countries made the transition to producing their own tradable capital goods, workshop copies apart? The few exceptions have already been mentioned. It must be supposed, with these exceptions, that producers in the low wage countries could not have obtained these higher rates of return from producing capital goods, which means that the assumption, that all countries have access to the same techniques of production, does not hold. That this assumption fails cannot be because of the physical and chemical processes of manufacturing, for, as argued in Chapter 1, they are well enough understood by those who devised them that they can be expected to be much the same in all countries, i.e. the effects of climate and terrain are known and can be avoided or are small enough to be ignored. It must, therefore, be because countries vary in the characteristics of their producers, workers and institutions on which the production of some goods depends. These

characteristics need not be unalterable and it is assumed here that they can, in any country, be so altered as to make the production of a good there like the production in any other country.

Whatever they be, therefore, these characteristics determine the international trade of the low wage countries with the high wage countries and those that matter most must be identified. The only practicable method is to see if, among the characteristics by which countries differ, there are any that seem to be linked to their trade. Conspicuous among them is that the high wage countries have more scientists and engineers relative to their populations than the low wage countries and their capacities for training them are also greater, and it is these countries that produce the technically more complex goods. Moreover, what were the low wage economies in the years immediately after the Second World War, first Japan, then Taiwan and South Korea and more recently Brazil, China and India, are the exceptions in that they have been producing technically complex goods and each has the capacity for training scientists and engineers in large numbers to the standards of the high wage countries.

The assumption that all countries have access to the same techniques of production must, then, be modified to allow for technical complexity. Goods differ in the technical and scientific knowledge needed to design them, to design the processes for making them and actually to make them. Much of the required knowledge can be transmitted through formal training, i.e. is taught in the education systems of countries, including the universities and institutions of that level, and is generally available, in print or some other medium to which all have access. Most such training is in the sciences and engineering and people trained in this way are referred to here as “trained workers”. Apart from such formally imparted knowledge is the knowledge of the untrained worker, such as the farmer or village weaver in a low wage country, who have specific knowledge essential to their production but acquired informally. Some knowledge may not be transmitted because it is ability acquired through experience, that is to say by performing the activity concerned, or because it is ‘proprietary’, meaning that some firms or individuals have exclusive rights to it and can withhold it from others. Goods that are technically complex are likely to be differentiated and it can be assumed that, if one version is tradable, all are.

The schema used here is straightforward; goods and versions of goods are assumed to be either “complex” or “simple”. Complex goods can only be produced with “trained workers”, people who have received higher education, especially scientists and engineers, and have the special knowledge for making those specific goods. Simple goods can be produced by workers without higher education. This dichotomy allows discussion of

technical progress, which can be expected to affect complex goods more than simple ones. In addition, it is assumed here that all tradable capital goods are complex.

An assumption of Chapter 1 that is kept throughout is that the high wage countries are the price setters for the goods they produce and the low wage countries for all other goods. Thus, the goods for which low wage countries are price setters at the start are primary products such as coffee, tea, bananas, vanilla, cloves and jute. Otherwise, these countries obtain the same income from producing a good as would a high wage country if the good were produced there and trade in that good and the inputs for making it were free.

What follows is an attempt to show with these assumptions that the explanation of why low wage countries so rarely became producers of tradable capital goods is that countries do not all use or do not all have access to the same production techniques, and that the reason for that is that the requisite knowledge is not equally available to all. First, all knowledge is assumed to be generally available, in which case a reason low wage countries cannot use the same production techniques as high wage countries is that they do not have the trained workers needed for making complex goods. Another reason can be added, namely that a firm's output and costs in producing a complex good depend on the experience its workers have gained producing that or other goods. It is the infant industry argument. The consequences for trade of each of these reasons are described using the schema of Chapter 1 and it is shown that neither gives an adequate explanation.

Second, it is assumed that not all knowledge is generally available. This allows consideration of R&D and technical progress and, through them, a cogent, plausible explanation can be given of what is observed. R&D also explains the phenomena described in the first part of this chapter. Trade is assumed to be free, for there are no other grounds for believing that the trade barriers of the high wage countries in the form of tariffs and import quotas prevent low wage countries from making capital goods.

Free Trade with All Knowledge Generally Available

When low wage countries have no trained workers: low wage goods

Beginning with the simplest case, low wage countries are assumed to have no trained workers; their firms can only produce simple goods, which are all consumption goods, with imported capital equipment. Using the schema of Chapter 1, they begin by producing cloth. First they displace the exports of high wage country cloth producers to low wage countries and

then, assuming the high wage countries do not set up trade barriers, they displace the production in the high wage countries for the domestic market until they produce all the cloth. After that they begin to produce some other consumption good and the process is repeated. At each such step competition reduces their rate of profit and the prices of all goods being produced in the low wage countries. Real wages rise in all countries as nominal wages stay constant.

If a low wage country reaches full employment while the high wage countries are still producing simple goods its nominal wage may start rising and its profit margins falling. That country may begin to produce goods that the countries with unemployment do not produce and to displace the high wage country exports of those goods to itself and other low wage countries, after which it may displace the production in the high wage countries for their own markets. If all the low wage countries reach full employment while the high wage countries are still producing simple goods, their nominal wages may begin to rise, though there is no assurance that they do or that they rise by the same amounts. To the extent that they do rise the prices of the goods made in these countries rise, too, while the same process of displacement of production goes on. There is no assurance either that real wages rise in the low wage countries when nominal wages rise since prices rise; they fall in those countries where the nominal wages rise too little and they fall in the high wage countries.

If the low wage countries take over all the production of simple goods from the high wage countries before they reach full employment and their nominal wages do not change, they bring the prices of these goods down by their mutual competition to yield a lower rate of profit and less income than could be obtained when the high wage countries still produced these goods. Since tradable capital goods are complex, firms in the low wage countries cannot mitigate the loss by producing their own capital equipment as described in Chapter 1; production for export is their substitute for a capital goods sector. In effect, the productivity of these economies is reduced; they must export more for the same imports, which is to say that, since to export is to save, they must save more to increase their capacity to produce an additional unit of income. They also obtain less income from their production than the high wage countries at that time. Goods whose prices are determined solely by competition among firms of low wage countries that cannot produce any capital goods that may be needed for making them are referred to here as "low wage goods". It can be supposed that the rate of profit in low wage countries settles at roughly that of the high wage countries.

The same results as in this second eventuality are brought about immediately when the high wage countries impose tariffs on the imports from

low wage countries that compete with their own production. Then, as long as the tariffs do not lower prices too far, the low wage countries continue to increase production of simple goods, but with the same effective loss of productivity. When the tariffs are high enough simple goods become low wage goods.

When low wage countries have trained workers

In the next case the low wage countries as a whole establish education systems that train scientists and engineers at the required levels so that their firms can produce complex goods under free trade. It can be assumed that these trained workers are normally paid less than workers with the same qualifications in high wage countries, though more than untrained workers in their own countries. Then, as these education systems expand, the production of some complex goods shifts from the high wage countries to them. High wage country firms either shift their own production to these countries or let the production of low wage country producers replace their own production for their own markets and their exports. They can, themselves, produce such goods in the low wage countries by investing directly in setting up production capacity in the low wage countries, either in subsidiaries they set up through FDI or in joint ventures with local firms, and employing local workers. Unemployment need not arise in the high wage countries; the rate at which their firms set up capacity in the low wage countries is limited by the numbers of trained workers and they provide some of the capital goods needed by these workers. But the saving rate may have to rise to allow the foreign as well as the domestic investment.

The process is the one described in Chapter 1, with the modification that the extent of production of complex goods in the low wage countries is limited by the numbers of trained workers available. While a complex good or version of a good is produced in both high and low wage countries the price is set by the former countries and the profit rate from making it in the latter is accordingly higher. Firms in low wage countries, including subsidiaries of foreign firms, increase their production of complex goods as the numbers of trained workers in these countries increase, starting with the goods that yield the highest rates of profit, whilst the production of simple goods yielding low profits continues alongside. Firms in high wage countries adapt their production accordingly. After the production in low wage countries of some goods has increased to the point that they are no longer produced in the high wage countries firms in low wage countries start producing other complex goods produced in the high wage countries at lower profit rates. The prices of the former goods fall to yield the same profit rates as the latter and are, in this way, still indirectly determined by the high wage countries.

The process continues as the numbers of trained workers in the low wage countries grows, until these countries reach full employment. During that time, if firms in low wage countries invest in the production of simple goods, for which the rate of profit is lower than for complex goods, the only reason is that all the trained workers are already employed. They may then bid up the pay of trained workers so that the profit rates from producing either type of good are equal, at which point the pay may even be higher than in the high wage countries, though that would depend on the amount by which the rate of profit on simple goods exceeds the rate of profit in the high wage countries and the proportion of workers employed in the production of complex goods who are untrained. Alternatively, they may invest little or not at all beyond replacement in producing simple goods or they may accept the difference in the profit rates and bide their time.

What happens to the prices of simple goods depends on how the expansion of education is spread across the low wage countries. If it is sufficiently concentrated in a group of countries with a small enough total population, only in these countries will all workers eventually be employed in the production of complex and untradable goods and none in producing simple tradables. Then the wages of trained and untrained workers in these countries can rise, whilst they remain low in the other low wage countries as long as they have unemployment, and simple goods stay low wage goods. Alternatively, the expansion of education is so spread over the low wage countries that, though some countries make complex goods, all make simple goods. If they eventually have full employment, the prices of simple goods start to rise and the wage of untrained workers, perhaps, too. Prices of simple goods must rise because there are presumably by then enough trained workers for firms producing simple goods to change to the production of complex goods and, with the barrier to competition between these two types of goods gone, the rates of profit on the former must rise. All prices are determined directly or indirectly by the high wage countries, which continue to produce some complex goods. The wage can rise as well, since unemployment, the barrier to that, is gone, and can, in principle, rise to the level of the high wage countries, with a corresponding fall in the rate of profit.

Whether or not wages of trained or untrained workers rise in low wage countries that have stopped producing simple goods and have full employment depends on the spread of education in other low wage countries. For instance, countries with big populations may expand their education systems fast and yet need a couple of generations to reach the level of the high wage countries. While this goes on, the rate of profit in any country where full employment has caused wage rates to rise will be lower. Such higher wages will be prevented if the firms of these countries, too,

invest abroad, though such investment may be prevented if not enough trained workers are available in the countries that have unemployment.

As long as the low wage countries train enough workers, though nominal wages there and in the high wage countries be constant, real incomes increase, mainly because of the growth of employment and profits from the production of complex goods. But in low wage countries the difference between the higher pay of trained workers and the wage of untrained workers can be affected by the pace of investment in the production of complex goods and by the numbers of people being trained in the natural sciences and engineering. Profit rates on the production of complex goods being higher, not only does total profit rise as the production of complex goods grows, but its share in the value of output rises too with the share of that production, though the part that accrues to foreign investment goes to the foreigners who own the investment.

Income in high wage countries is increased by FDI in low wage countries since it yields higher rates of profit than domestic production. Beyond that, it cannot be said *a priori* whether the income increases or decreases with trade. Faster increase of the number of trained workers in the low wage countries and enough production in the high wage countries of machines to employ them in making complex goods causes the production of these goods in the high wage countries to be displaced faster. On the one hand, equipment there has to be scrapped faster and this raises costs. On the other hand, the additional machines the high wage countries export may have a larger profit component, so that, for the same workforce, the value of the exports is greater and more can be imported from the low wage countries when trade is balanced. It is also possible that not enough machines will be imported into the low wage countries to employ all the trained workers, which means that the saving by the low wage countries in the form of more exports and in the high wage countries in the form of more FDI or loans to low wage country firms is too low. Structural unemployment in the high wage countries cannot be ruled out *a priori* either. For instance, trade may balance but the number of workers in the high wage countries displaced by imports from low wage countries does not exactly match the numbers taken up directly and indirectly by the production of machines for export. If those displaced are more numerous than those taken up, there is unemployment and if they are fewer, trade does not balance.

Experience

Defined for present purposes as ability acquired through performing various activities, experience can be considered to comprise several forms of knowledge. One is knowledge in the same sense as a person is said to know how to play the piano or to ride a bicycle. In these the experience is gained

through practice and following example, and may be accompanied or preceded by the formal teaching of generally available knowledge, as with the musician, carpenter or ship's captain. Apprenticeship commonly comprises practice, learning by example and formal lessons, and, explicitly or not, forms part of the training in many activities and professions. Experience is also one way the architect, the civil engineer, the various designers of motor car exteriors, fashion clothing or furniture and the head cook of a hotel improve their abilities. Experience can here be taken to include a form of knowledge referred to as 'tacit'⁷², which has not been written down but is imparted by word of mouth and example and acquired from practice.

Experience can be specific to the product, the equipment or both. The same capital equipment may yield more output per period with the same number of workers or require fewer workers for the same output, the gain being greater the more of the good that has been produced by those workers. But the experience may have to be acquired anew with each new design of the product. In some branches of production relations of this sort have been found empirically to be systematic and stable. Arrow refers to what may have been the first estimate of this kind, Wright's estimate in 1936 of the man hours needed for the production of airframes. Here the learning seems to have been specific to the product more than to the equipment. Adam Smith's pin makers seem also to illustrate this type of experience; they seem to have gained manual dexterity on the job and could not have transferred it to making other goods. Arrow also refers to the steady rise in output per head of the Horndal steel works, which occurred without new investment and was, therefore, specific to the equipment as well as to the product. Learning curves have become routine both in the production of new versions of complex goods and the planning of new plant.

Such specific experience results in output per worker being greater in firms with greater total previous output, other things being equal, and the need for it can hinder firms starting the production of goods they have not produced before and can, in principle, lead to monopolies. Experience must be supposed to have diminishing returns; there are technically determined limits as to how much output can be increased or labour and other inputs can be reduced using the same equipment. During World War II more estimates of 'learning curves' were made for aircraft production and yielded the S-curve according to which, with each new type of aircraft, the man hours per aircraft declined with the number already produced, slowly at first, then accelerating and then decelerating to an asymptote.

⁷² Kim, "The Dynamics of Technology Development: Lessons from the Korean Economy", 92; Cyhn, *Technology Transfer and International Production*.

Experience can also be diffuse in the sense that, over time, industry creates an industrial culture. Workers and managers who are part of that culture adapt to new tasks and equipment more easily than their counterparts in countries where there has been little industry and the simultaneous existence of different industries can result in improvements in individual activities without there being obvious links.

Attempts to devise mathematical models of experience have to cope with processes in which the past affects the present, which already makes them intractable without the additional complication of international trade. Arrow seems to have been the first to attempt a model, though he conflates experience with technical progress. Experience in his model is 'learning by doing'; as the amount of past production grows the result is not more output or less labour for the same machines, but new, more productive machines. Experience results in new machines that require fewer workers, but does not change the numbers of workers the old machines require. Machines and consumption consist of the same substance, but a later machine is different to an earlier one and produces the same amount of the good with less labour. There are no marginal products and the real wage can vary. Some such simplifications are, however, a practical necessity for a deterministic, mathematical model.

Such learning by doing can conceivably result in trade because of an absolute advantage. It may be possible to adapt Arrow's model to trade between two countries by assuming that the experience gained by producing machines in one country is not transferred to the other country. One country, then, produces all the machines, up to the limit of using all its workforce for that, and exchanges machines for consumption goods made in the other country. Both countries can produce the consumption good with the same machines using the same amount of labour, but one is more efficient at producing machines. This sets limits within which the wage is exogenously given for each country.

Lack of trained workers or experience as inadequate explanations

Education is more complex than industrial production, but training scientists and engineers is investment similar to the production of capital goods. If education is profitable in the high wage countries, which the private schools and universities show it is, it should be profitable in the low wage countries. In theory, it can be more profitable than in high wage countries if the trained workers coming out of it are paid at levels not too far below the levels of those countries.

A sign of the complexity of education is that, in practice, profit does not ensure enough is provided to meet all the need, for, in all countries, much of

it, usually the greater part, is provided by the state. As a profit making business education has been spreading in many low wage countries, but it is restricted to those whose families can pay for it. If, therefore, education, especially primary and secondary schooling, is lacking in a low wage country, it is because the state lacks the means or the desire to provide more. There have always been cases where those who govern the country are insufficiently concerned about the welfare of the people, but, even when they are concerned, their budget revenues rarely meet the state's various needs. This is, to some extent, an indication that the political leaders, rightly or wrongly, think the economic return on education is not high enough to warrant more budgetary expenditure as compared to other needs.

Profit is evidently not a motive to meet the standards of the high wage countries either. Usually the demand for education in the low wage countries is enough to fill the schools and universities, but, except for some in the rapidly industrialising economies of East Asia and India, these institutions do not train scientists and engineers to the levels of the high wage countries. Establishing education institutions of high enough standards is difficult and slow, but that, after several decades and with these exceptions, virtually no universities of low wage countries are considered to be on a par with the established universities of the high wage countries implies that reaching such standards is not profitable.

A priori, if the level of training and the pay of a scientist or engineer in a low wage country were the same as in a high wage country, the profitability or economic return on education would be correspondingly higher. If the possibilities of providing such training are lacking, there is a presumption that it is not expected that such training would result in sufficiently high pay. If having the suitably trained workers sufficed for firms in low wage countries to produce complex goods competitively, it would have been possible to pay these workers enough for the training to be profitable. Training could have started on a small scale for one or two activities, but would have spread. The resulting production could not have been stopped by the high wage countries, even if they imposed tariffs on imports like those on cloth in Chapter 1, for the low wage country firms could turn to producing the complex goods, especially the capital goods that their countries had been importing. Nothing could have stopped the low wage countries from becoming self-sufficient if trade with the high wage countries did not yield higher rates of profit.

It is in theory possible that every complex good requires so much experience to produce that a low wage country firm beginning to produce one would be unable to compete with established high wage country producers. But there are reasons to doubt that this is the reason why low

wage countries, with the exceptions mentioned, have not become makers of complex goods. One is that there is such a variety of complex goods that it would be expected that some would not require so much experience as to offset the low wage. Another is that lack of experience seems not to be a handicap when high wage country firms set up manufacturing in low wage countries. One example of many is the production capacity for making motor cars for export in Mexico. A third reason is that, again with the same exceptions, experience gained under protection has not made low wage country firms set up to produce complex goods competitive. Many of these firms closed as trade was liberalised and those that survive still need protection. India's industrial sector, for instance, developed under protection, but is not an exporter of complex goods. Pakistan has been producing cars for the domestic market with imported and locally made components for some decades, but the production remains protected by the country's highest tariffs.

Free Trade with R&D

R&D, proprietary knowledge and product differentiation

It was argued in the foregoing that two phenomena of international trade need to be explained. One is product differentiation, which, itself, explains why countries buy and sell the same goods and why the law of one price does not hold. The second is the inability of low wage countries, with a few exceptions, to produce their own tradable capital goods and complex goods in general. Neither shortages of trained workers nor the need for experience is an adequate explanation of the second. In what follows it is argued that both phenomena are explained by technical progress that is the result of deliberate spending, i.e. R&D. It suffices for a realistic and plausible description of much of trade, so from here on all technical progress is assumed to be the result of R&D.

No representation of technical progress and R&D by mathematical formulae can be realistic enough to be useful; any discussion has to be qualitative and verbal. Technical progress includes improvements of consumption goods and the invention of new ones, and there is no objective method for quantifying these. This is ignored in the standard discussions of technical progress, in which output is a known quantity and technical progress is identified with the part of the increase of output not attributable to increased use of factors. Such technical progress is, thus, the unexplained residual derived from some assumption about the relation between factors and output, e.g. the form of a production function. Capital goods, too, are improved and new ones invented and they cannot be quantified either. So, neither all the inputs nor all the outputs of the standard estimates of technical progress can be

quantified. Similar remarks apply to the models mathematical devised to bring R&D in as the source of technical progress, the endogenous growth models, though these, in addition to simplifying by quantifying inputs and outputs, simplify R&D as a quantity of labour or final output. Their conclusions do not and could not be expected to bear any resemblance to reality. These points are treated systematically in the appendix.

R&D has costs and firms incur them because it generates knowledge that can give an advantage over competitors, who can be prevented from using that knowledge. Before machinery and factory production replaced artisans producers in the finer crafts could keep knowledge from rivals by organising in guilds and the like because it was mostly acquired by word of mouth, example and practice. These crafts had to be learnt in workshops regulated by guilds, which set the standards and prevented the knowledge from being imparted outside their purview. A household could spin wool and small towns had weavers for weaving cloth of wool or linen for the common people, but the well-to-do wore the finer woollen and silk cloth woven by the workshops of the guilds and beyond the skill of the ordinary weaver. Secrecy was the other main means of keeping knowledge from rivals, for guilds or workshops often had processes that rivals did not know. As machines manned by unskilled workers replaced the craftsmen and produced new goods, knowledge as skill gave way to knowledge of process. But it was often knowledge that a rival could discover by examining the product or by spying and only in the twentieth century did patent and copyright laws come to protect it internationally. Ohlin's example of the linen industries of Northern Ireland and Southern Scotland may seem to have been a case of restriction of knowledge to a group of firms without apparently using secrecy or the formal rules of guilds, but competitors elsewhere would probably have acquired equivalent knowledge if they had had no better investment alternatives.

Nowadays a firm owns the knowledge its R&D generates, it is "proprietary knowledge" and remains proprietary as long as the patents and copyrights are in force and others are prevented from using the results or must pay for the right to use them. Secrecy is still normal, but in the following, knowledge that is not generally available is assumed to be protected by patents, which expire after a fixed time. R&D is mainly done by private firms, though some may be undertaken by state institutions, including universities, and some of the costs of the R&D of the private sector can be borne by the state and some of the costs incurred by state institutions by private firms. Goods, then, are of three kinds; complex goods that are directly affected by R&D, either because they, themselves, or their processes of manufacture are frequently improved by R&D, the other complex goods and simple goods. Goods made by processes that are permanently secret are not discussed. If all complex goods are affected by R&D, there are only two kinds of goods.

Producers are assumed to be in competition, which implies that every good of the first kind can be made in several versions, for every version is protected by patents and trademarks. It is also assumed that R&D results, occasionally or continually, in “improved” versions and that firms price these new versions so as to obtain as much profit on their investment, including the cost of the R&D, as possible. Improvements in production processes occur also and can be attributed to improved versions of capital equipment. As new versions are brought out, buyers pay less for old versions and the profit margins from making them, net of depreciation, decline. Eventually the old versions are displaced. Producers anticipate this and calculate their rates of profit on the investment in productive capacity and R&D from the appropriately discounted values of the profits and investments. Firms compete through R&D, then, by bringing out new versions of goods that displace older versions.

As a rule, a firm that has not been producing a particular good will be unable to generate proprietary knowledge as advanced as that of firms that have been producing improved versions and have kept their proprietary knowledge up to date with R&D. Normally a firm’s R&D generates new proprietary knowledge using proprietary knowledge it already has, so that a firm that has not produced the good before and does not have a suitable stock of proprietary knowledge of its own needs to spend on R&D for several years just to be able to make versions of the good that would have been competitive at the time it began the R&D, i.e. would have yielded the same rate of profit as obtained from other activities given the prices then of other goods. In the meantime the other firms that have been making the good have gone on spending on R&D and improving their versions. Unless it obtains the up to date proprietary knowledge of the more advanced firms, the newcomer firm remains backward compared to them.

Buying proprietary knowledge

Whether or not high wage country firms sell their proprietary knowledge and at what price depend on what effect the addition to the competition will have. When the producers of a good are so many that their competition approximates the perfect competition of textbooks, no firm having so great a share of the market as to need to take account of how its actions affect other firms, a firm’s sale of its knowledge to another has no noticeable effect on the amount of competition. But in the contrary case a firm that sells its proprietary knowledge to a firm that has not been making that good has a new competitor to take into account, in particular one that produces versions of the good close to its own.

Competition through R&D is likely to result in the number of firms being too small for firms to ignore how others might react to their actions; if the

number of firms actually doing R&D is so large that competition is close to perfect, it diminishes until only a few are left. When competition is close to perfect there must be a market for proprietary knowledge and, so, there must be firms that buy it rather than do their own R&D. When only one firm is willing to sell its proprietary knowledge to a firm that has not been doing its own R&D while keeping the right to use that knowledge itself, it can, in theory, expect the buying firm to pay up to the amount that yields the same rate of return as it would get in other activities. Whatever the amount, for the selling firm it is income additional to its normal profit from production. When every firm is willing to sell its proprietary knowledge, the price each gets for it is, in theory, lower than what it would have been if the firm were alone in selling, and so must be the prices of its versions of the good. Firms that sell their knowledge receive income both from that sale and from the production of their versions of the good, whilst competition reduces their rates of profit to the same as in other activities, and those that buy save the costs of R&D to make the same rate of return.

Comparing the situation in which firms cannot sell their proprietary knowledge to that in which they can and assuming the same number of firms in both situations, fewer firms spend on R&D in the latter and total expenditure is lower and spread over the same number of firms, but, if the expenditure by each of those that does R&D is the same in both situations, the lower expenditure does not mean slower improvement of the goods produced by these firms, but fewer versions. It is implicit in this kind of competition that firms do their R&D in parallel and duplicate each other's work in the sense of obtaining similar knowledge without violating patents. As long as the number of ways a good or its production may be improved is not smaller than the number of firms needed for an approximation to perfect competition, this competition can continue.

Two mechanisms make it likely that the number of firms producing a good is reduced to the point that each firm takes in to account how the others might react to what it does. One operates when the number of ways a good or its production can be improved is smaller than the number of firms doing R&D, so that those who apply for their patents first grow at the expense of the others, who lose the right to use the knowledge involved, unless they pay for it. This is likely to occur when the number of firms is great enough for competition to be close to perfect. Then the supposition of perfect competition, that firms are much alike in the quality of their R&D has to be abandoned; firms that are better in this respect obtain the patents and the others close. The other mechanism is that R&D gives increasing returns to scale because its cost is independent of the scale of production of the version of the good that results; firms with greater output can spread the costs of R&D over more units. This can occur with complex goods with many

characteristics that can be improved through R&D, for the firms producing them may judge it necessary to spend on a wide array of improvements.

Because of these two mechanisms, the rule is that big firms that have been producing a good for some time have an advantage over small firms and newcomers. A newcomer that has to begin producing its first versions of a complex good using its own R&D may not be able to find designs or methods of making the good that have not already been patented by firms already producing the good. And if it is small with a relatively small outlay on R&D it is unlikely to acquire up to date knowledge and catch up with a big firm on its own, let alone to do so profitably. While incurring the costs of its R&D at the start the newcomer receives no income from it, while that of the firm already producing the good is up to date and, as it goes along, results in new versions of goods that yield roughly the same profit rates as do other investments. More spending on R&D can be expected to result in improved versions more quickly, so big firms are likely to increase their superiority. And if the newcomer wishes to have an R&D program equal to that of a big firm, it must be able to expand its production accordingly, which means both that it must have the productive capacity ready in time and that its versions of the good should sell well enough. For the firm with the bigger output a similar increase of capacity is proportionately smaller and is further reduced to the extent that existing capacity can be adapted from producing the old version of the good to producing the new one. Hence, the big firm can amortise its investment over a longer time, with correspondingly lower costs per unit.

This does not mean that smaller firms in high wage countries cannot continue or even grow to displace big firms. Partly it is a matter of appearances; a firm may only seem small compared to firms in other activities, but may be big in its own activities because they are specialised. Such firms are not exceptions to the rule and can provide a large part, perhaps the greater part, of the output, outside retailing and farming, of a high wage country. True exceptions to the rule occur when small firms invent new goods, descry markets for known goods or discover ways of improving their products that have been overlooked by the bigger firms or have become possible because of scientific developments outside these firms. Sometimes big firms neglect their R&D or misdirect it and decline as a result. Such things occur repeatedly, but the rule remains, in straightforward competition the bigger firm has the advantage.

Competition between R&D and low wages

Low wage country firms cannot be expected to be able to do R&D comparable to that of high wage country firms. Nevertheless, they may be able to compete; using generally available knowledge they may, because of

the low wages they pay, be able to make versions that are backward but cheap enough to have some buyers. If they can compete like this, these firms receive less income per worker than do the high wage country firms making the same goods, but their profit rates are at least equal to that from the production of simple goods. As the patents of the high wage country firms expire and more knowledge becomes generally available, the low wage country firms can improve on the versions they produce.

It cannot be said *a priori* how much production of this kind will occur. It is also practically impossible to tell empirically. There have been no surveys to find out and, though it can be obvious if a version of a good is up to date or not, often, particularly with capital equipment, only an expert can judge. Besides, a backward version of a complex good can be partly more advanced than versions made only with generally available knowledge. It can include components that were, themselves, bought from a firm generating advanced proprietary knowledge or were made with knowledge bought from such a firm and which are more advanced than its components made with generally available knowledge, perhaps even up to date. Or the low wage country firm may, in time, start its own R&D and generate knowledge that is less advanced than that of the long established high wage country firms, but better than generally available knowledge. It is enough to know that low wage country firms can, in principle, sell backward versions of complex goods cheaply in competition with the advanced version of the high wage country firms and that they can do so on a large scale.

Yet, even on a large scale, this type of competition may not suffice for a low wage country to reach full employment, despite having the trained workers needed. For one, the prices of such backward versions of goods are determined by buyers' preferences and by the up to date versions made by the high wage country firms, both of which are beyond the influence of the low wage country producers. Thus the price of a version produced by a low wage country firm may, because of a new version made by a high wage country firm, drop so low that it stops being profitable. Just the lack of access to the most advanced knowledge prevents the firm from being able to foretell what the next versions will be like and how they will affect the demand for its own versions. Such demand is precarious; it may permit large scale production at one moment and then suddenly fall. These countries still have unemployment when all the production of simple goods and of the complex goods on which R&D has stopped has been transferred to them. In this outcome the tradables the high wage countries produce are only those to which they have devoted R&D, whereas the low wage countries, though they produce some backward versions of these goods, produce the other complex goods, which have become low wage goods, with due allowance for the higher pay of trained workers, and for the rest simple and untradable goods.

Competition can also reduce the margin by which the pay of trained workers in the low wage countries exceeds the wage of untrained workers.

It follows that a low wage country firm cannot obtain the same income in relation to its investment from production under free trade as do the firms of high wage countries, though that does not preclude it from obtaining more than it does from producing low wage goods. To obtain the same income as a high wage country firm, it must acquire the knowledge for producing complex goods that fetch correspondingly high prices, i.e. knowledge equivalent to the knowledge of the former. Since it cannot be expected to generate such knowledge from its own R&D, i.e. progress faster than do the high wage country firms, it must find ways of inducing these firms to impart it. It can try to buy the knowledge, in which case its income from production is that much less. But it may not even be able to buy it at a price that allows it to make a profit. One reason is that, if trade and capital flows are free, the payment must be high enough to be more attractive to the high wage country firm than the alternatives of selling the good in the low wage country or of setting up a subsidiary to produce it there. Such payments become continuous, for, as new versions are produced by the R&D of the high wage country firm, new knowledge has to be bought. Should the low wage country firm stop buying new knowledge, its versions of the good become more outdated and the income they yield diminishes with time.

A second reason that the cost of buying knowledge may be too high for an adequate profit is that the number of high wage country firms producing the good is small enough that each takes into account the possibility that selling its proprietary knowledge can result in a new competitor, one, moreover, whose versions of the good in question are likely to be close to its own. If the buying firm has adequate R&D capacity of its own it can try to generate its own knowledge and it may, in due course, become able to design versions of its own that differ enough from those of the selling firm not to violate patents or infringe on brands and enable it, therefore, to compete as an equally advanced producer. Because of this, there may be no price at which the firm selling the knowledge considers itself to be compensated for the addition to its competitors and the buying firm gets profits as good as it can get otherwise. If the buyer does not have adequate R&D capacity of its own or there is no scope for designing new versions or improving production without violating patents, it cannot produce up to date versions of the good at all; it cannot make its own versions, whereas the seller does not allow it to produce versions too like its own unless the buyer sells to it, in effect subcontracts.

High wage country firms, then, sell their proprietary knowledge to low wage country firms if they believe that the effects of the addition to the

competition resulting from the sale will be small enough to be offset by the price or some other gain. One case is that of high wage country firms that make a particular good judging that the scope for improvements is too small to justify the R&D involved, for example, because R&D has for some time yielded little improvement. They then stop R&D and competition eventually brings the price down. Eventually low wage country firms become able to compete using generally available knowledge as proprietary knowledge ages, and in time proprietary knowledge becomes generally available. Any high wage country firm owning such proprietary knowledge will be willing to sell it knowing that it will in time lose its value, that competing firms will be willing to sell theirs too and that the competition of low wage country firms will come about anyway. Consequently, it will be willing to sell it at a price that low wage country firms will pay to begin production before the good becomes so cheap as to be a low wage good. Alternatively, a high wage country firm can avoid an outright sale by entering into partnership with a low wage country firm, transmitting the proprietary knowledge in return for a share of the profit or some form of subcontracting, such as the right to some of the output.

There are other circumstances under which a high wage country firm will sell its proprietary knowledge to a low wage country firm, as when a firm decides to abandon a particular activity and no longer cares whether there is competition or not, or when a firm sells subsidiaries making losses in the belief that they will not make the buyer a competitor of importance. For low wage country firms these are opportunities for acquiring proprietary knowledge of high wage country firms. In either case other high wage country firms may try to prevent the sale to a low wage country firm by buying the proprietary knowledge or subsidiary themselves, knowing that it may eventually result in a competitor with advanced knowledge and low wages. If the costs, which may include debt incurred from making losses, are too high, or if they believe that the low wage country firm will not, through its own R&D, become a serious rival, they will refrain from such a pre-emptive purchase. Thus, a high wage country firm may sell an unprofitable motor car maker to a low wage country firm, because it judges that the buying firm does not have the R&D capacity to keep up in the future, whereas the buying firm, if its market is protected against import competition, judges that the proprietary knowledge thus acquired will enable it to make cars in its own country that will be more profitable there. Alternatively, a low wage country firm may buy outdated knowledge and obtain adequate profits from cheap versions that are less advanced than those of the selling firm, or buy knowledge that is up to date but only suited to making versions that compete little with the versions of the selling firm. Such arrangements can be repeated regularly.

Buying proprietary knowledge may also be unprofitable because the good, itself, is new and the decline in the number of firms making it has not ended. In that case, buying the knowledge for making the good is risky because the firm selling the knowledge may not survive “consolidation” and may be willing to sell because it, itself, doubts it will survive. Then the knowledge loses some of its value. Examples of goods that were made by many firms when they were new and later came to be produced by a few with big outlays on R&D are motor cars and personal computers. In the early years of the motor car there were several dozen manufacturers in each of the industrial economies of Europe and scores in the US. Now there are just a few independent makers in the whole world and the number has diminished from decade to decade since World War II. Similarly personal computers, invented in the 1970s, were produced in the 1980s by a variety of firms, big and small, many of which have given up or disappeared.

Self-reliance

The upshot is that, under free trade, income from investment in low wage countries falls, if it is not already low, because simple goods become low wage goods and complex goods for which there is no proprietary knowledge may also do so in due course. But low wage country firms cannot, as a rule, expect to generate their own proprietary knowledge through their own R&D and produce goods that yield the same income as the goods made by high wage country firms nor expect these firms to sell them the proprietary knowledge that will let them do so. They may at some time be able to produce backward versions of complex goods using generally available knowledge and their own R&D and obtain better returns on their investment than with other complex goods, but the scope for such production may be small and the returns of these investments over time are determined by changes that are unpredictable and beyond their influence. If a low wage country is to have some assurance of being able to obtain more income from its investment than available from producing low wage goods it must find other ways of investing than those just described. What follows describes briefly some alternatives.

If trade is free and the authorities of a low wage country judge that their country’s firms can only expect to remain producers of goods yielding low income because that is the most that versions of goods made with generally available knowledge can yield and because the cost of up to date proprietary knowledge is too high, they may also observe that the living standards of the bulk of the population in the high wage countries fifty years ago were higher than they are for the bulk of the population in the low wage countries now, which implies that using generally available knowledge allows higher incomes than does the production of low wage goods, for patents do not last

fifty years. That is to say, if the production methods and designs of goods of fifty years ago could be applied in the low wage countries now with the rate of profit of the high wage countries of then, a real wage would result that would allow consumption per head of that time's goods and saving at least as high as they were in the high wage countries then. Workers would consider themselves better off than they are now because they would be able to buy some goods that, though backward, they are too poor to buy now. An economy can be viable without being internationally competitive.

Nevertheless, replicating goods and processes of an industrial system of long enough ago for the proprietary knowledge of the time to have become generally available is not possible under free trade. It would put out of date goods in competition with imported up to date goods. A motor car, aeroplane, dishwasher or some industrial machine made with the technical knowledge of twenty years ago would not nowadays be bought at prices that covered the costs of production of an efficient low wage producer when up to date versions are available, even if safety regulations and the price of fuel have not changed. Much the same can be said of most manufactures that are regularly improved with R&D; television sets, telephones, many types of capital equipment, medical apparatus, etc.. In general, a version of a complex good that yielded the income of the high wage countries' wages and rate of profit of the time it was produced and used in the high wage countries cannot be sold years later at a price that yields the same wages and profit rate. If enough time has elapsed for the proprietary knowledge to have become generally available knowledge, its price may have to be so low as not to yield the wages and the profits low wage countries get from producing simple goods. How low the price must be is a partial indication of the rate of technical progress in the design of that good; other things being equal, a slower rate allows a higher price. Low wage countries that produce goods using generally available knowledge, when high wage country firms produce up to date versions because of their R&D, obtain low rates of profit because prices change.

It cannot be argued that what lowers the prices that the low wage countries would receive for complex goods they would be able to make is not the new versions, but new methods of production brought about by R&D that these countries cannot use. This can be argued for individual goods but not generally. If, for instance, it were true of all consumption goods, it would imply that the standard of living in the high wage countries at the time when the latest knowledge that is now generally available was new was lower than in the low wage countries now because production costs were higher.

A type of change that can, *a priori*, prevent low wage country producers from selling older versions of goods in high wage countries, but will be

ignored here, is regulatory change, in particular, regulations governing the effects on the environment and safety, such as the prohibition of the use of lead in the fuel for cars and or rules for the packaging and paint of consumption goods. It is possible that satisfying such regulations requires proprietary knowledge, but new standards are not normally imposed if the producers in the high wage countries who can meet them are the exception. So, it is assumed that they can be met using generally available knowledge.

Those authorities of a low wage country who follow this train of reasoning have one certain way of raising the income, i.e. wages and profit, yielded by investment above that obtained from producing low wage goods, namely to protect its production against competition from imports to the point that domestic producers can produce and sell complex goods using generally available knowledge. It has been done, India being an example from its independence to the 1990s. An illustration is given by Wolf, who wrote in 1982 regarding Indian industry, "Although the older products were not necessarily inefficient, without large price discounts they were often not attractive in world markets. ... the designs of even the best manufacturers' trucks were considered out of date."⁷³

This differs from the standard infant industry argument in that protection is permanent. One form of the argument of Hamilton and List is that protection allows industries in countries where they are new to gain the experience they need to compete with the established industries of other countries. Accordingly, protection is a temporary necessity to be removed once enough experience has been gained. But, as argued above, domestic producers may not be able to compete with established foreign producers, who offer more advanced versions of the same complex goods, because there are difference in the technical possibilities available to each country. Protection is needed permanently, for the domestic producers cannot be supposed eventually to become competitive with the established producers, though, as patents expire and generally available knowledge increases, incomes rise.

Autarky, therefore, allows investments to yield more income than does free trade in a low wage country that does not have to import primary goods like food and raw materials. If the country can export some primary products, the investments may yield still more income if they include imported capital equipment that is more up to date than what can be made with generally available knowledge, provided that any special knowledge needed to operate the equipment is supplied with it and is promptly learned. To this extent the country's producers make complex consumption goods designed with generally available knowledge and use imported equipment in some

⁷³ Wolf, *India's Exports*, 66.

parts or the whole of the production processes. Assuming technically advanced equipment that can be used for such production to exist, it must not be priced so high as to offset the gain in output or the reduction of labour costs. For instance, equipment used in the high wage country may be operated with so few workers and cost so much, that it is less profitable to use in the low wage country than equipment made there. Between these two types of equipment there are suitable designs of capital goods that were used earlier by producers in high wage countries, when wages were lower, and there may be technically advanced capital goods designed for low wage countries.

For producers in the low wage country the criterion for importing a capital good to make a particular good is straightforward, namely, whether or not, given domestic prices and the nominal wage, it increases the return on the investment, which happens either because the imported capital good costs less for the same output per worker than ones made domestically or because it yields more output per worker. In contrast, if the country has to export simple goods to import such capital goods and, perhaps, food and raw materials, the choice is more complicated, because exports require some form of subsidies. Its higher income means its real wage or its return on investment is higher than in the low wage countries that have free trade and, therefore, the prices of its simple goods are higher, so that it must either subsidise its exports directly, or it must have subsidies to keep the nominal wage, but not the real wage, lower than in low wage countries that trade freely. For the present it is assumed that, if the real wage of one low wage country is higher than that of another, its nominal wage is not lower.

In practice, countries are no longer free to raise trade barriers; they are almost always bound by international agreements, including those of the World Trade Organisation (WTO), of which the purpose is to foster free trade, and the conditions for policy loans by the World Bank and International Monetary Fund (IMF). In many low wage countries the economists and some of the authorities have adopted the orthodox belief that free trade is economically efficient and in all of them they know they must reach agreement about their trade protection with the high wage countries through the WTO, for these countries are their main markets for exports and sources of foreign aid. If they had high tariffs, they may have bound them higher than others for a while and, as in India, continue to produce out of date versions of goods while gaining time for acquiring the knowledge for making up to date versions. It depended on how much the countries relied on exports and foreign aid and on their authorities' negotiating skill.

For a low wage country, therefore, the main problem of economic development is that of finding tradable goods they can produce that yield more

income than do low wage tradable goods without protection against imports. Failing that and leaving aside special cases of natural or historical endowments, such as minerals or tourist attractions, the country is confined to competing with other low wage countries in activities that generate low incomes, often with high capital costs and rarely with much scope for growth. Orthodox economists do not think of the problem in these terms but believe that the choice of production activities is best left to market mechanisms undistorted by efforts of government officials to guide investment. They specifically advise against “picking winners” since, as experience shows, the activities chosen by government officials usually turn out not to be competitive, and are, therefore, relieved of any obligation to give specific advice. Undistorted implies the least impediments to international trade, with due allowance for infant industries, an objective these countries are under obligation to achieve through agreements as members of the WTO and with the multilateral and bilateral donors. Nowadays government officials and their advisers are little inclined to object since they have usually been schooled in the orthodoxy and know, as well, that their economies’ needs to import and to service foreign debt preclude breaches of these agreements that can provoke retaliation with cessation of foreign financing and obstacles to exports.

Cooperation: subcontracting

If trade is free, therefore, the firms of a low wage country can only enter into activities that yield more income in relation to investment than do low wage goods if high wage country firms impart proprietary knowledge to them. High wage country firms have two reasons for not transmitting such knowledge; one is that the firms receiving it may become competitors and the other that they cannot be sure that the knowledge will not be obtained by others without their agreement. But they also have reasons to transfer production to low wage country firms. One is to lower wage costs. A second is to increase output when full employment at home is a constraint. Thirdly, they may need to circumvent trade restrictions. Trade is assumed here to be free, but, in practice, high wage countries sometimes restrict or threaten to restrict imports from other high wage countries, so this eventuality may need to be allowed for.

Often a good that is the result of proprietary knowledge includes components or its production has steps that require only generally available knowledge and which can, therefore, be transferred to low wage country firms that have the trained workers, if needed. It can be done by setting up a subsidiary or by subcontracting to a low wage country firm, but is no more than an extension of the production of simple and complex goods already described. Since it entails no proprietary knowledge, competition keeps it the production of low wage goods. Should any proprietary knowledge have

to be provide, it will be confined to workers transferred from the parent firm or it will be of a kind that cannot be of use to an existing rival firm or for starting a new one. Workers of subcontractors or of subsidiaries may receive training from the high wage country firms, but they are not given access to proprietary knowledge.

But a point can come at which a firm of a high wage country has transferred all activities that do not require proprietary knowledge that it can to low wage countries and further growth requires that some proprietary knowledge be transferred as well. Then there is cooperation between the high wage country firm and the low wage country partner, which may be an independent subcontractor, a subsidiary or a joint venture of the high wage country firm with a local firm. The high wage country firm provides its partner proprietary knowledge and training as needed. Both firms can obtain higher profits, for there is a range for the prices of the components that the low wage country firm produces; they can be so low that they yield profit rates no higher than what the firm would obtain otherwise, or so high that the high wage country firm cannot increase its profits or lower its prices. Both, therefore, benefit by agreeing to prices between these extremes. Thus the low wage country firm supplements its partner's workforce and increases its own output with higher profit rates, whilst its trained workers increase their technical knowledge and, perhaps, get paid more.

Both firms are likely to have an interest in expanding the scope of such collaboration and, therefore, in making it stable and stability requires that each firm believe that the other judges lasting collaboration to be in its own interest. If the low wage country firm is independent and judges that its subcontracting arrangement will not lead to enough expansion of output, diversification of products or transfers to it of knowledge, it can be expected to try to find other high wage country firms to collaborate with. It may have little to lose if the high wage country firm has limited its subcontracting to a few items with little transfer of knowledge and not much gain in profit. If it so wishes, the high wage country firm can reassure the low wage country firm by widening the collaboration, giving the latter more components to produce, especially ones that require more of its proprietary knowledge. It can even subcontract some R&D to the extent that the other firm has the capacity. It thus makes an investment, in the form of a transfer of its proprietary knowledge, intended to give a return over several periods, and it can expect that, later, providing new knowledge generated by its R&D and the associated training to the firm with which it is already collaborating will, unless that firm has some specific drawback, cost less than would starting collaboration with another firm.

The high wage country firm has to be sure that collaboration will not result in a new competitor and that the knowledge it transmits will not be passed on without its consent. It can prevent the former simply by restricting the collaboration and the knowledge it transmits, but then it removes the condition for stability. To prevent the latter it needs a deterrent, recourse to the law or ending the collaboration, though neither is convincing unless the appropriate conditions are met. Recourse to the law depends on the effectiveness of the legal system and on having enough evidence, whilst ending the collaboration is convincing to the extent that the low wage country firm depends on the collaboration and the high wage country firm does not. Either firm can reduce its dependence by collaborating with other firms, preferably on the same products, but must balance that against the lessening of the scope of each case of collaboration and, therefore, of its stability, whilst the high wage country firm also depends on the collaboration as a restraint on the low wage country firm passing on the knowledge it has.

A better way of making collaboration stable is for the low wage country firm not to be independent, which is to say that the high wage country firm acquires a share in ownership of the low wage country firm or sets up a joint venture with it, in either case with a position in the management. By this the high wage country firm confirms its intention that the collaboration last, the condition for it to expand in scale and variety. It probably still transfers knowledge through licence agreements, i.e. sells it to the low wage country firm, despite its part ownership, though, perhaps, at a lower price. Apart from yielding income, the licence agreement specifies if and how the knowledge can be passed on, whilst the high wage country firm's position in the management allows it to ensure that the agreement is respected.

Collaboration of this kind can evolve to the point that the low wage country firm accounts for enough of the production of a good for the good to be marked as made in that firm's country, the brand name remaining that of the high wage country firm, what is termed "original equipment manufacturing" (OEM). This does not mean that the good is entirely produced in the low wage country. In particular, the high wage country firm will supply the technically most advanced or complex components if they cannot be produced by the low wage country firm or if it does not want to transmit the relevant knowledge, this latter perhaps because the low wage country firm has over time acquired the R&D capacity and employs the trained workers that can make it a competitor.

Both the firms and the authorities of a low wage country may decide that subcontracting as described here, in which the production is to the specifications of the high wage country firms that have the proprietary knowledge for producing the final goods, may not suffice for the long run.

Such subcontracting has two drawbacks. One is the uncertainty from dependence on the proprietary knowledge of firms of other countries. Sometimes the costs and other disadvantages to the high wage country firm of starting a new arrangement with another low wage country firm give little protection to an established collaborative arrangement. Should, for example, the high wage country firm already have similar arrangements with other low wage country firms, shifting production from one to another at little cost may be simple. Another reason that one low wage country firm can be replaced by another comes from new knowledge and the consequent changes of goods and production processes. The high wage country firm is likely to know of impending changes, which may even be partly or wholly caused by its own R&D, and can prepare to adapt, whereas the low wage country firm, though it may be aware of the new knowledge, does not have the proprietary knowledge that would allow it to anticipate how its subcontracting arrangement may be affected. Change can also reduce the advantage it has over other low wage country firms of the training and knowledge it received from the high wage country firm.

The second drawback is that there is no reason that the subcontracting should grow over the long run at a proportional rate as is expected of a growing economy. Production can, in principle, generate the capacity for future production, which is the reason for the frequent assumption of proportional growth. But, when it consists of this kind of subcontracting, it depends on the amount of production that the high wage country firms wish to pass on to the low wage country firms and on the numbers of trained workers in the firms of other low wage countries. Even if subcontracting yields high growth rates at the start, the authorities of an economy with a big population may think it imprudent to assume that it will yield similar growth rates in the long run. They will then try to find ways of acquiring proprietary knowledge comparable to that of the high wage country firms by inducing these firms to impart more of their knowledge or by investing in R&D in the hope of catching up.

Alternative: inducing the transfer of proprietary knowledge

The authorities must, therefore, be able to offer modes of cooperation that make transmitting proprietary knowledge more profitable for high wage country firms than the alternatives. That means, firstly, that the good in question must be produced in the country. This can be brought about by trade barriers and it is assumed for the present that the authorities have the power to regulate trade barriers and to prevent foreign investment. They can, then, raise the profit to the foreign firm from making the good in the country and, if tariffs and competition from other firms do not lower the profit margin from selling the good in the country when it is made elsewhere, they can

restrict the amount imported. They thus oblige the foreign firm to choose between, on the one hand, lower profits or even being altogether excluded from the country and, on the other, reaching agreement on a mode of cooperation by which it transmits proprietary knowledge, and the firm makes its choice taking into account the protection such cooperation gives its proprietary knowledge and the other determinants of its profitability, in particular, the size of the population, its income and the prospects of growth.

Cooperation is then likely to be a joint venture. It has to last several periods, for it can be supposed that the investment by the foreign firm in the country does not yield an adequate return in just one period and that the firm wants to keep enough control over the proprietary knowledge it transmits to be able prevent it from being passed to others while it has value. Moreover, both the authorities and the local firms involved want regular transmission of new proprietary knowledge as it is generated by the high wage country firm's R&D and, as is often necessary, the adaptation of that knowledge to local requirements. Since the foreign and the local firms involved have an investment in it, a joint venture gives all a motive for cooperation of a kind that the authorities also want.

If the foreign firm already has an assembly operation in the country, the authorities can impose local content requirements, i.e. require that some portion of the components that are imported be replaced by local production. It applies in particular to subsidiaries of foreign firms assembling goods, like cars, air conditioners and television sets, from imported components, which, because of import barriers, can be profitable even if the foreign exchange cost of the imported components exceeds that of the final good.

Economies with bigger populations and faster growing incomes can obtain more favourable terms, for, as with R&D, the proprietary knowledge transmitted is independent of the amount of output and the high wage country firm, therefore, more willing to impart its knowledge for the prospect of more sales. In particular, a bigger country can have more foreign firms competing to supply their proprietary knowledge for producing the same good, possibly with lower trade barriers and lower prices. In a small country, other things being equal, a foreign firm may require higher protection against competition for a joint venture, even to the point of becoming a monopolist. If, as is usually the case, the authorities cannot promise that, they will not be able to obtain proprietary knowledge through cooperation of this sort.

Besides joint ventures, there are few other methods of obtaining proprietary knowledge, unless a high wage country firm is willing to sell its proprietary knowledge at a price that yields an acceptable rate of profit to the buying firm. One possibility is that the low wage country firm acquires enough control over a high wage country firm by buying some or all of its

shares to be able to use its proprietary knowledge and R&D, though this requires that the low wage country firm already be big enough or have ample financing available, perhaps with the help of the state. It is easier if the firm being acquired is in difficulty and would otherwise close or lay off many workers, though, if the reason for the difficulty is that its proprietary knowledge is inferior to that of its competitors, the acquiring firm has to rely on a combination of investment in R&D and production of cheaper versions of the good in question by paying lower wages. Just the acquired proprietary knowledge may suffice to give the firm an advantage over competitors in its home country. Paying lower wages means some production, at least, will be in the low wage country in part, and will not prevent workers being laid off. Whether or not the prospect of a low wage country firm becoming a new competitor and, perhaps, catching up incites other high wage country firms to prevent the acquisition, either by competing for the firm being acquired or by persuading their authorities to interfere in some way, depends on, among other things, where the acquiring firm intends to sell its output, how much of the financing the acquiring firm receives from its state and how these authorities balance their free market principles against the unemployment prevented and interests of their countries' firms.

Whatever the process by which the proprietary knowledge is transmitted, the authorities and firms of the low wage country must have between them the technical competence to negotiate and supervise formal agreements. How well they judge the knowledge acquired depends on the quality of government scientific institutions, the knowledge of the firms and their trained workers and on the advice of foreign experts, but they are unlikely to be able to specify what the most advanced knowledge is or to know enough about the R&D of big firms that produce many varieties of goods to make useful comparisons. Leaving aside for the moment international agreements, such as treaties under the WTO, in the early stages an agreement might specify the state's obligations to the foreign firm regarding matters such as trade barriers, foreign exchange and capital controls, bank financing, payment of profits and might go so far as to include agreements with other foreign firms in the production of the same goods. In due course, as more and more agreements are reached, the scope for variation in the state's obligations, i.e. the state's control, is reduced to permission to invest, both for the foreign and domestic firms, and financing.

3. ECONOMIC GROWTH SINCE WORLD WAR II

International Trade and Payment Arrangements

The new trade arrangements

Nothing has been said so far about the system that actually governs international trade and payments. What is now the system that prevails over the whole world evolved in stages from organisations and rules that were deliberately designed at the end of the Second World War. Designing such a system was unprecedented, but unavoidable. The system that had evolved through the 18th and 19th centuries, culminating in harmonious trade among most countries and the gold standard in all industrial countries as well as most others that were not parts of empires, had been brought to an end by the First World War. It was natural after that war to imagine that, once the gold standard was restored, the old system of trade would return by itself, for what had functioned so well before had not been deliberately designed and seemed, therefore, to be a natural state of affairs. Besides, an international agreement would not have been possible; enmities were too strong, the US would not participate in such an undertaking and Europe now included several new countries that had been parts of the Austro-Hungarian and Russian empires. What ensued was a period of political and economic turmoil including the Great Depression and ending in war.

This period had shown that a satisfactory system of trade and payments would not come about by itself, but had to be designed and constructed deliberately. It was to be a system to govern trade, exchange rates and payments between all countries outside the communist bloc, though, because of the circumstances, the people who designed it were mainly a group from the US, led by Harry Dexter White, and a group from the UK led by John Maynard Keynes. Between them they established the principles that were to govern the new system and the requisite organisations and mechanisms. They disagreed on several matters, but were in broad agreement on what had gone wrong and what they wished to bring about. Their ideal was freedom of movement of goods, services, capital and ownership, though not of people, between countries and in pursuing their ideal they established an institutional framework that motivated their countries to persist so well that, half a century later, the ideal was to a great extent realised for themselves and most of the rest of the world.

Political circumstances were also different; where the only powers of consequence after World War I had been the Western allies, now there was the USSR, both militarily powerful and in occupation of all Europe east of the middle of Germany and not on the Mediterranean Sea. Moreover, communist parties in several countries of West Europe had large followings

with memories of the unemployment and poverty of the years between the two wars and fears of more of the same. Political leaders of the US, hostile to communism as they were, could not, as they had done after the Versailles Treaty, withdraw from Europe, but were obliged to collaborate with the many Western European leaders alarmed by communism in allowing a high degree of political freedom to their peoples whilst bringing about enough prosperity to allay the desire for communism. Proposals for turning Austria and Germany into rural economies were rejected and demands for reparations were limited to specific categories of people who had been persecuted by the Nazis, in particular Jews. Reparations for Poland and Russia were mainly settled by transfers of territory. Western European countries also set up systems of social protection and allowed wage earners enough bargaining power for them to share in the growth of income to such an extent that the share of wages came to seem an economic constant. There followed decades of peace among the industrial economies and a period of economic prosperity for Western Europe, North America and Japan that surpassed anything that had gone before.

The high wage countries reduced barriers to trade among themselves gradually to avoid economic disruption and to be mutually advantageous, though by the 1980s they had lowered them enough for this trade to count as free, at least for present purposes. They gave the low wage countries roughly the same conditions for their exports of complex goods as they gave themselves, but restricted their own imports of these countries' simple goods. In establishing the General Agreement on Tariffs and Trade, replaced in 1995 by the World Trade Organisation, they set up the mechanism to foster free trade. Trade negotiations under the aegis of these organisations consisted of bargaining, or 'reciprocity', with the condition that the terms a country accorded to any party should be as favourable as those accorded to any other party in every respect, the principle of the Most Favoured Nation (MFN). High wage countries could reduce their protection gradually with the prospect of exports compensating for any losses of employment and productive capacity.

These rules do not, by themselves, prevent the countries that have or will have bigger markets from discriminating systematically against the others, yet any such discrimination against the smaller countries of West Europe was minimal. Countries with small markets exporting a good produced in all the countries with big markets have little to offer for a bargain. But the authorities of the bigger countries of the West were, until the 1980s, too concerned about the communist parties and the control of almost all Eastern Europe by communist governments for there to be systematic discrimination within West Europe. Apart from the right wing dictatorships of Portugal and Spain, Western Europe consisted of electoral democracies where trade discrimination could have caused resentment and divisiveness that were better avoided.

High wage countries also have less motivation to protect their production of complex goods against the competition of other high wage countries than they have for protection against the competition of simple goods from low wage countries, both because the high wage countries are usually bigger markets and their threat of retaliation more serious and because of the diversity of complex goods. Similar complex goods are likely to come from several countries, big and small, and, whereas discriminating against producers in the big ones can be expected to provoke retaliation, discriminating against producers in small countries often only affects minor sources of supply. Retaliation is specifically provided for by the GATT and the WTO, which set the rules for it. Complex goods can also be so diverse in their variety or their components that the affected industries of the high wage countries are too disparate to be likely to organise opposition to imports. At the other extreme, when the production of one good is dominated by the firms of just one or two countries, the other countries, being importers of that good, have no immediate interest in opposing imports from a new source. Thus, the German firms making cameras, which had dominated until then, could not have organised protection by the other high wage countries when Japanese firms began to displace their exports in the mid-1950s. In a few years Japanese cameras dominated the market as much as the German ones had done. In other respects, nevertheless, Japanese firms were often discriminated against and later they circumvented more discrimination by transferring some of their production to other countries, including the high wage countries to which they exported. By then Japan had become a high wage country and, like other high wage countries, its firms were also transferring production to low wage countries. Consequently, much of high wage country imports from low wage countries consisted of production by subsidiaries of high wage country firms or goods ordered by them, so that the main support for protection, though with little effect, was that of the workers, who were losing jobs because of the transfers of production and whose ability to negotiate higher wages was being reduced.

Industries producing simple goods, like cotton textiles, are confined to a small range of products and, so, opposition to imports was easy to organise. Against this, low wage countries had little bargaining power; their markets were small because of the low incomes and because the authorities of the big countries, like Brazil and India, did not reduce their own protection. Consequently, the high wage countries were not only slower to reduce barriers against imports of simple manufactures from these countries that competed with their own production, but, despite the MFN principle, discriminated against them.

Their protection against cotton textile imports from low wage countries, which started in the mid-1950s, was discriminatory, for their protection against imports from other high wage countries was more lenient. Being

against the spirit and the letter of the GATT, it was formalised under the Short Term Cotton Textile Arrangement in 1956 with the assurance that this was a temporary measure to allow the high wage countries to adapt. In 1960 the Arrangement was replaced by the Long Term Arrangement, with the assurance that it would not be extended to fibres other than cotton. The low wage countries had to accept it as a condition for being allowed to take part in the Kennedy Round of negotiations of 1962 to liberalise trade. The Multi-Fibre Agreement (MFA) was adopted in 1974 and lasted with modifications until 2005, whereupon the quotas ended, though some high tariffs remained. Similar trade barriers have been placed by the high wage countries on most of the other manufactured exports of low wage countries, garments, leather goods, jute products, etc..

Another part of the process of discrimination has been the creation of trade blocs, which have reduced barriers within the blocs, though they sometimes eliminated preferences the exports of some low wage countries had been receiving as members of former European empires. The European Common Market comprising six countries developed out of the European Coal and Steel Community (Montanunion) of 1951 and the Treaty of Rome of 1957, whilst seven other European countries joined in the European Free Trade Area. Out of these came the European Union. Countries as diverse as Pakistan and New Zealand lost any trade advantages they had had with the UK and were placed in the same position as all other non-European countries. Other trade blocs have been the North American Free Trade Area (NAFTA) and several set up by groups of low wage countries in Latin America, the Caribbean and East Africa, though, since the imports of most importance to these countries came from outside the blocs, they have had little effect. Other departures from the MFN principle that have become commonplace since the 1970s are voluntary export restraints and bilateral agreements.

Globalisation mitigated the effects of these discriminatory trading arrangements on the income of the low wage countries that were limited to the production of simple goods by extending the range of simple goods. Among other things, it caused the transfer by high wage country firms of the production of simple components or of simple steps of production processes to low wage countries, as described in the schema above, which seems to have begun in 1961, when two high wage country firms making electronic goods began to have some simple operations carried out by unskilled workers in Hong-Kong and Taiwan.⁷⁴ After that it spread to other simple goods and processes and to other countries. But for that, the growth of exports of the low wage countries and, hence, the growth of income would have been slower, for the reduction of high wage countries' barriers against

⁷⁴ Wade, *Governing the Market.*, 94.

the imports of these goods was drawn out over more than half a century, time enough for the high wage countries to divest themselves of the production of what were the simple goods at the start.

Most such production, especially in the early stages, used capacity set up as FDI by high wage country firms in low wage countries for the purpose of producing goods to their specifications. Such FDI provided the country receiving it employment and income in foreign exchange without costing investment in productive capacity, except as might be needed for providing water, power, local transport, security and other services, much of which also brought in income. So, most low wage countries' authorities tried to attract such investment. Among the inducements they offered were exemption from most or all taxes, simple import and export procedures, special government agencies to take care of their concerns and, occasionally, ready made structures for housing the machinery and offices. A common arrangement of this kind was the export processing zone (EPZ), which provided foreign firms most or all such advantages and accorded them a specific legal status. Much of globalisation consisted of the spread of production in this way and it accounted for a large part of the economic growth of countries like Malaysia, Mexico with its *maquiladoras*, Thailand, Tunisia and Turkey, which did not, at the start, have the ability to produce complex goods or do any R&D.

Theory and political constraints in the low wage countries

The schema that has been given here provides an explanation as to why low wage countries that have continued to produce only simple goods have stayed poor and why the ones whose incomes have grown fastest are those that have been producing complex goods. But it cannot explain why there were so many of the former and so few of the latter. Nor can it, by itself, explain why the actual course of events was not as orderly as the schema, itself. Explanation of all this must to some extent come from outside it and must consist of a combination of three different components, namely, the circumstances of international trade, the beliefs of the authorities of the low wage countries and the extent to which these authorities could determine what happened.

Taking circumstances into account mainly means allowing for the degree of protection and for the beliefs about protection of the authorities of both high and low wage countries. Hence the schema, which has been presented on the assumption of free trade, has to be elaborated accordingly. Both the changes in the degree of protection of the high wage countries and the slowness as regards the simple goods exported by the low wage countries have to be taken into account.

Similarly, account must be taken of the beliefs of the authorities of the low wage country and how they have changed. Up to the 1970s nearly every

independent low wage country's authorities thought that domestic industries had to be protected against competition from imports. But, their countries depended on economic aid from the international institutions, such as the World Bank, the regional development banks and the IMF, and were continually influenced by them. All these institutions promulgated the orthodox belief, that a country that could not much influence its terms of trade did best by accepting world prices, i.e. by removing trade barriers, even without reciprocity. It has been the theory taught in the universities, especially in the high wage country, with little indication from the 1980s on of alternatives, so that the authorities of the low wage country, especially the higher level officials, have come more and more to be composed of officials who have been educated to believe in it and to believe that their countries' balance of payments and debt problems have been the results of inefficiency caused by protection. These are the officials who are now deemed qualified to negotiate the agreements to get the financing their countries need and who have gradually replaced the others.

Besides, in many low wage countries the authorities changed to a belief in free trade only because they had no alternative theory or schema to refer to and they usually did so with qualifications. Apart from the orthodox argument for free trade, all they had was the infant industry argument in the form it commonly takes nowadays, namely that a low wage country firm starting to make a complex good not already produced in the country is, at first, inefficient in terms of use of current inputs per unit of output and of actual output in relation to the labour and productive equipment employed, but becomes more efficient with experience.

What distinguished the authorities and firms of the first low wage economies to produce complex goods from the others was that they had understood early that the relevant knowledge had to be acquired. It took them time to find out for themselves how to do it and, having neither theory nor well formulated practice to refer to, they made, as to be expected, mistakes and had failures. They, nevertheless, became proficient at identifying and obtaining knowledge that served their purpose and eventually built up their own capacity for generating knowledge. The schema given earlier is a summary of some of what they found out for themselves and the methods they devised.

Having started early, they had a choice of methods that has since narrowed as trade has been liberalised and some methods have been banned under international agreements. For some time, when their own countries protected their industries, the authorities and economists of the high wage countries refrained from requiring the low wage countries to liberalise their trade, however much they recommended it. This changed in the 1980s and

1990s. Application of Intellectual Property Rights also became stricter and acquiring proprietary knowledge without permission of the owner became harder; subsidies and cross subsidies within firms, which had enabled infant industries of complex goods to become sufficiently advanced and efficient to compete with imports and, in time, as exports, were forbidden, as were performance targets, by which the authorities induced firms to export for low returns or even losses. Individual countries could not choose to avoid GATT and WTO negotiations to reduce ordinary protection against competition from imports without being discriminated against. In addition, trade liberalisation became a regular condition for economic aid from the international institutions, especially for the many low wage countries that went through balance of payments crises and needed both financing from these institutions and debt relief, though this did not necessarily entail reciprocity on the part of the high wage countries.

In some low wage countries the conditions, social, political and environmental, or domestic and foreign interests were such that authorities would not have been able to take the steps for firms to begin producing complex goods, even if they had wanted to. In countries like those of the Sahel the extreme environmental conditions would probably have required more investment in infrastructure and education than the countries could finance. In Pakistan owners of big agricultural properties have had the political power to prevent a tax on agricultural incomes, which made the financing of an adequate educational system practically impossible. Several countries had tribal and social systems with groups at the head whose power and income would have been reduced by modernisation. To put it differently, that the authorities of some low wage countries should have taken the steps necessary for the production of complex goods may have been the result of special circumstances, of specific social, historical and political conditions, which are beyond the scope of the present work.

Since all low wage countries had, at the start, to import their complex goods, especially capital equipment, from the high wage country, their authorities had to decide what their countries should produce for export then and later and which of the complex goods they imported should be made domestically. Some believed that the protection of the high wage countries would so hinder exports and that the terms of trade would so deteriorate that their countries should produce some, at least, of the capital goods they needed themselves.

That this reasoning was logically consistent was shown by Raj and Sen in 1961 by a model in which the value of exports was assumed to be constant and in which a consumption good was assumed to be made using a capital good with a current input, both of which were made using a second type of

capital good, which could also be used to make its own type.⁷⁵ Raj and Sen concluded that in the long run the highest rate of consumption came from limiting imports to capital goods of the second type. Atkinson reformulated the model as an optimisation and reached much the same conclusions.⁷⁶

Their assumption that the protection of the high wage countries would hinder the growth of exports of simple manufactures of low wage countries and that the terms of trade of these countries exporting these goods and primary products would worsen over time was reasonable, but it did not follow that the same would be true for complex goods that these countries might export. If they had assumed that the country could export the current input and the two capital goods at the prices of the high wage countries as long as they were produced in these countries, they would have come to the conclusion of Chapter 1. It would have been equivalent to what Japan and, later, Taiwan and South Korea found out, that, as the high wage countries lowered the barriers against imports of complex goods from one another, they lowered them for other countries as well.

The training of scientists and engineers in the low wage countries

One prerequisite, therefore, for low wage countries changing from producing simple goods for low incomes to making complex goods that would allow them higher incomes, was that their education and training systems be adequate for training scientists and engineers of the required levels and in the numbers needed.

None of the low wage countries had the necessary capacity for training scientists and engineers at the end of the Second World War, except for Japan, then a low wage country though it had reached the technical levels of high wage countries in some respects and in a few, such as lens making, had gone higher. It is beyond the scope of this work to explain why this was so, but the extent of the dependence of such training capacity on the state in the high wage countries leads to the conclusion that private individuals and institutions in these countries did not provide enough to meet the existing demand, let alone to meet future needs. In the low wage countries, where production requiring trained workers had not yet started, therefore, it was not to be expected that such capacity would be created as profit making businesses or as private non-profit or charitable institutions. That the authorities did not think it necessary to do more themselves seems to have been in large part because they and their foreign advisors regarded it as relatively unimportant, even a luxury, and believed that poverty was better reduced by primary education,

⁷⁵ Raj and Sen, "Alternative Patterns of Growth under Conditions of Stagnant Export Earnings."

⁷⁶ Atkinson, "Import Strategy and Growth under Conditions of Stagnant Export Earnings."

though they might also be deterred by the fear that high unemployment among the educated could cause political troubles.⁷⁷

In some cases, however, as independence in India in 1947, revolution in China with migration of the old regime to Taiwan in 1949 and a military coup in Korea in 1961, the new authorities set up such capacity because they did judge it necessary. Over the following decades they set up the capacity to train people to the levels of the good universities of the high wage countries. That high proportions of the undergraduates in the high wage country universities nowadays come from these economies is the effect both of the extent to which the growth of the production of complex goods has caused the demand for such education to increase faster than the domestic capacity to provide it and of the growth of secondary education, and both are the effects, on the one hand, of the ability of governments, firms and households to finance such education and, on the other, of the judgement of these and of the students that it leads to suitable jobs. At post-graduate levels the proportions are also high, presumably because the universities of the low wage economies still lack the capacity to provide greater specialisation at the most advanced levels. In particular, at post-doctoral levels training and gaining experience depend on research, for which the financing is usually provided by the industries and governments of the high wage countries to the institutions of their own countries.

The low wage countries make two groups; the few that have built up education systems that train enough scientists and engineers to the standards of the high wage countries to meet the needs of their industries and the great majority that have not. Several of the latter have begun to invest in setting up some capacity for such training, occasionally to the point of being able to provide training in some branches to the levels of good high wage country universities, but, compared to the first group, their capacity is still small in relation to the population.

Having scientists and engineers trained in the high wage countries, which all low wage countries do to some extent, does not make up for a lack of domestic training capacity. It requires a system for financing large numbers of people who lack the means themselves, which those authorities who consider training at university as not of great importance are unlikely to set up. Consequently, students from low wage countries in high wage country universities come mainly from families who can pay for their children's studies and have no personal motive to change this. For them and their families there is the additional motive of having qualifications that make it easier to emigrate to the high wage countries, whilst for these countries they are workers trained at little or no cost to themselves. Moreover, the

⁷⁷ Laïdi, *Enquete sur la Banque mondiale*, 66.

number of people who qualify for good high wage country universities depends on the quality of the secondary schools, of which only a few are likely to be of the level needed, especially if the authorities judge preparing students for higher education not to be of sufficient value.

Some More Successful Economies

Japan

When the Second World War ended the only low wage country to have the trained workers needed for producing complex goods was Japan. Its firms also had technical knowledge at levels close to those of the firms of the western countries for producing a variety of goods that were needed for industrial production, like steel, various kinds of machinery, industrial chemicals, etc.. In some respects, notably optical lenses, they seem to have been ahead. In principle, they had no knowledge that was not generally available, for acquisition of new knowledge had stopped just before the War and, as with Austria and Germany, the victorious occupying powers, in this case the US, eliminated the legal protection of Japanese proprietary knowledge. If some proprietary knowledge of Japanese firms did remain proprietary and more advanced than what was generally available, it was because it was overlooked by the occupying power.

Once the worst shortages of the aftermath of the War were over, the authorities had to decide on the degree to which they would direct the economy and determine or influence the activities of firms. Industrialisation after the Meiji Restoration of 1868 had been the result of a deliberate effort to learn the scientific and engineering knowledge of the Western countries, an effort that was directed by the state. It was to be expected that Japan's political leaders, government officials and firms would believe that they would again have to obtain from abroad the knowledge they did not have but needed. They had to choose also how much emphasis to give to what was referred to as "heavy industry and chemicals", as opposed to letting demand be determined by consumers and exports. By "heavy industry and chemicals" was meant the capital and intermediate goods needed for industrial production, including plant, machinery, ships, steel, non-ferrous metals, inorganic industrial chemicals and petrochemicals for synthetic fibres and plastics. Such industries were often referred to as "heavy" at the time, for the term, though vague, did describe most of them through the nineteenth and early twentieth centuries.

Eventually the authorities fostered both industries producing the means of production and acquisition of the knowledge needed to reach the levels of the firms of the Western countries, especially the US, for the goods they intended that the economy should produce. They had the powers to oblige

or persuade firms to conform to their intentions, for they, in particular the Ministry of International Trade and Industry (MITI), controlled foreign exchange and financing, and their permits were needed for many things, such as investing and buying the proprietary knowledge of a foreign firm. They could alter taxes without opposition from politicians and, as in almost all countries at the time, they determined import duties. They also gave several inducements to firms to invest in R&D with the intention of reaching the levels of the leading Western firms, including direct subsidies of R&D, tax advantages and organisation of coordination with other firms and institutions. Moreover, they set up the organisations for discussing regularly with the firms long term plans, immediate problems and how to deal with them.⁷⁸

Firms already began to acquire knowledge more advanced than what they had under the occupation, mostly by buying it through licensing and paying royalties. One of the reasons for the occupation authorities' approval of the Foreign Investment Law of 1950 was that it made sure that the foreign exchange would be available for royalty payments, though the Japanese authorities considered the law a means of confining the acquisition of proprietary knowledge to licensing and royalties, without letting foreign firms acquire ownership or management positions.⁷⁹ In 1963 the need for permits for such acquisitions for less than \$30,000 was removed, provided they allowed no foreign participation in management.⁸⁰ The law was also used by the authorities to extract better terms than the firms could obtain by themselves.⁸¹ Japanese firms' contracts to buy proprietary knowledge from foreign firms, "technology imports", had started by 1950 and were matched by contracts for technical and consultancy services. Each averaged around 100 a year until 1956, after which the former increased rapidly.

That Japanese firms could obtain most of the proprietary knowledge they wanted from firms whose competitors they were to become has three possible explanations. One is that there was more competition among Western firms then than, say, three decades later, by when R&D had become concentrated in a few relatively big firms. Western firms often refused to sell their proprietary knowledge, but usually there were others in the same activities that did not. On occasion firms agreed to transmit knowledge in exchange for permission to produce in Japan, and, because of the pace of the growth of the economy and its exports, even accepted to limit sales there, as happened with major computer makers.⁸² Another explanation is that the R&D capabilities of the

⁷⁸ Johnson, *MITI and the Japanese Miracle*.

⁷⁹ Johnson, 217.

⁸⁰ Krause and Sekiguchi, "Japan and the World Economy," 454.

⁸¹ Ozawa, *Japan's Technological Challenge to the West, 1950-1974.*, 545.

⁸² Johnson, *MITI and the Japanese Miracle.*, 247; Komiya, Okuno, and Suzumura, *Industrial Policy of Japan*, 180.

firms selling proprietary knowledge were so much better than what could be expected of the Japanese firms, that the selling firms did not foresee a risk. The third is that some of the firms selling their knowledge judged that R&D was not going to yield worthwhile technical progress and that their proprietary knowledge would soon be generally available.

When Japan began again to industrialise trade barriers were used to restrict imports of complex consumption and capital goods to those the authorities judged were needed but could not be made in the country or whose advanced design made them especially useful. At first the imports of these goods and raw materials, as well as the royalties for the proprietary knowledge its firms had bought, were mainly paid for by exports of simple goods, notably textiles. In 1950 textile exports, including raw silk, were alone 3.5 times the exports of metal and metal products, machinery and rolling stock and the ratio was still 1.7 in 1954. Japanese textiles, though cheap compared to those made in Europe or the US, can be assumed to have fetched good prices when compared to the prices other low wage countries received in the 1960s, for in the early 1950s textile exports by low wage countries had not reached the quantities that later provoked the formal protective arrangements of the high wage countries. Pakistan's textile industry had hardly begun. Nevertheless, Japanese textiles exporters already had to accept voluntary export restraints by 1953. These restricted the amounts exported but let the authorities organise exporters so as to prevent their mutual competition from reducing prices. They did, however, show what was in the offing.

At the start, therefore, Japan's post-war industrialisation consisted of investment in the production of capital goods using generally available knowledge and protecting its industries from the competition of more advanced versions of these goods made in the Western countries, primarily the US. Since it had its trained workers and since its firms had, at least, the knowledge that was generally available, the industrialisation would have continued in the same way, with Japan gradually exporting more complex goods of less advanced design than those of the high wage countries, probably mostly to other low wage countries, even if its firms and authorities had not acquired proprietary knowledge from foreign firms. As long as there were workers seeking employment in industry and as long as the prices of its exports did not fall too much relative to the prices of its raw material imports, its industrial growth for a given saving rate would have been faster than it could have been before the War because the generally available knowledge was more advanced.

The scale of acquisition of advanced knowledge from abroad resulted in a sustained growth of industry that was faster still and well beyond what had occurred before. It also resulted in exports of complex goods soon exceeding

the exports of simple goods. From the results, it seems that the judgement of the authorities as to which complex goods should be imported and which kinds of proprietary knowledge should be bought was on the whole sound. According to the MITI indices, the production of capital goods almost tripled from 1955 to 1960 and production of durable consumer goods increased more than fivefold.⁸³ Manufactures of other categories, consumer non-durables, construction materials and intermediate goods, did not grow at those rates, but still fast enough for the index of production of industry as a whole, comprising “public utilities, mining and manufacturing” to double and more in the same time. Textiles continued to be produced, but became progressively more complex and of higher quality as wages rose, whilst the production of simpler, lower quality textiles was taken over by other low wage countries. Much the same was true for the next ten years. As given by the MITI indices, production of capital goods and consumer durables rose fivefold from 1960 to early 1970 and industry as a whole more than tripled. For the same reasons, but with less direction by the authorities, the European countries whose economies had been similarly damaged by the War also had faster economic growth rates than they had ever had, though lower than Japan’s, and it was common to speak of the “German economic miracle”.

On occasion Japanese firms started exporting goods they had not exported before and displaced Western firms making those goods because they had developed related industries, that is to say because they had a variety of industries. Cameras is an example. Before and during the War Japanese firms had, by acquiring the knowledge from abroad and through their own R&D, learnt to make optical lenses that were at least as good as the best lenses made in the West. When they learnt to make the mechanisms for cameras in the 1950s, they displaced the German firms that had been the main makers of high quality cameras. Another example is that of the quartz watch. Watches were clockwork mechanisms in the 1950s and virtually all the higher quality makes were produced by Swiss firms, which also did the R&D. These firms knew that a quartz mechanism was likely to displace clockwork and tried to develop one, but it was the Japanese firms that first developed one commercially. The advantage of the Japanese firms was that Japan had an electronics industry, which Switzerland did not, and the quartz mechanism is electronic. Japan’s authorities and firms had decided early, while the economy was still limited to radios and the beginnings of television, that electronics was an industry in which they needed to reach Western levels. Since then cameras have become electronic devices, too.

A third example is that of motor cars. Japanese motor cars were produced under virtually total protection for decades, during which time

⁸³ The indices are 100 for 1960, 34.3 for capital goods and 18.9 for consumer durables in 1955. MITI.

investment in R&D allowed the makers to catch up with the western producers. Again, they were not far behind at the start and became equally advanced in the production of several components by the time they became major exporters in the 1960s. Motor cars were also a good of which backward versions could be sold cheaply in low wage countries and, since Japanese cars were solidly engineered, they displaced most Western makes there. Soon they were displacing Western makes in their home markets.

In due course the Japanese authorities had to give up their powers to influence firms directly, though the process of doing so began after Japan had become a fast growing, broad based industrial economy with fast growing exports of manufactures. They were conforming to a trend set by the Western countries towards free trade and open markets for goods, capital and services, other than labour, one that had to be accepted by all countries that wished not to be penalised in their trade with these countries. It was also a consequence of the growth of Japan's economy and exports. Japan had become a high wage country and deviations from free market practices that were tolerated while it was poor and perhaps threatened by communism were tolerable no more.

It joined the World Bank and the IMF in 1952 and freed current account transactions on its balance of payments in accordance with Article 8 of the IMF statutes in 1964. Japan became party to the GATT in 1955 and undertook to remove all export subsidies and foreign exchange budget allocations so as to reach Article 11 status, which it did by 1963. But exporters still got preferential credits and, from 1964 on, were given a variety of tax advantages for depreciation and reserves of various kinds. In 1964 Japan joined the Organisation for Economic Cooperation and Development (OECD), with obligations to accept the organisation's rules regarding competition and state interference, but was allowed to keep restrictions on capital movements. Nevertheless, it was criticised by the OECD in 1968 because the authorities repeatedly and successfully interfered in negotiations for purchases of proprietary knowledge by Japanese firms from foreign firms to the advantage of the former.⁸⁴ It was in 1968 that the first steps were taken to remove the restrictions of the Foreign Capital Law of 1950, though initially the activities in which foreign investment was allowed freely were either of no interest to foreign firms or did not interest Japanese firms. In 1978, the same year as the US and Switzerland, Japan ratified the Patent Cooperation Treaty. It reduced its trade barriers a little during the Kennedy Round (1964–72) and more during the Tokyo Round (1972–79), when the ratio of receipts from import duties to the value of imported dutiable goods fell from around 15 per cent to about 6 per cent. This made no difference to its trade balance and the criticisms of its trade barriers by the Western high wage countries, which had begun in the 1970s,

⁸⁴ Ozawa, *Japan's Technological Challenge to the West, 1950-1974.*, 54–55.

did not stop. Unlike the other high wage countries, Japan did not use quotas allowed by the MFA to protect its textile industry, and its tariffs on textiles other than silk were low.⁸⁵ The process of conforming to free markets went on, but Japan as a high wage country is beyond the scope of this discussion.

Taiwan

Taiwan and South Korea are the two economies that began after the War with low wages, whose firms had no proprietary knowledge of their own and produced few complex goods, and yet began early to increase the variety and quantity of complex goods they produced and eventually to generate their own knowledge. They had several things in common. Both had been occupied by Japan for forty to fifty years until the end of the War and had been used as extensions of the Japanese economy, not only for primary products, but to supply industrial, especially intermediate goods, as well. Naturally, almost all the trained workers had been Japanese. Despite the destruction, both had at the end of the War more industrial capacity and better infrastructure in relation to the sizes of their populations than all but a few low wage countries. Both had been cut out of larger countries and were considered by the Western high wage countries to be threatened by communist neighbours, and in each the period of rapid industrial growth was started by the deliberate actions of a new government, in Taiwan that of the Kuomintang leaders who came there when the communists gained control of China's mainland in 1949, and in South Korea that of the military government that deposed Syngman Rhee in 1961.

Not only was Taiwan more industrialised than most low wage countries, its industries were more diverse, including chemical fertilisers, petroleum refining, aluminium and copper refining and building of small ships. It seems to have had some trained workers from the Japanese occupation and more arrived from the mainland. They were enough for industrial production in 1952–3 to be as high as it had been before or during the War. The new authorities did not have a past in which they had systematically acquired technical knowledge from the West as a prerequisite for industrialisation, as did the authorities in Japan, but, like the Japanese, soon after taking control they showed that they wanted the economy to be able to produce complex goods and to acquire the necessary knowledge.

As with the Japanese authorities of the time, they had the means to oblige or persuade firms to conform to their intentions and to discuss with firms what could be done and how.⁸⁶ All foreign exchange had to be sold to them at the rates they fixed, which allowed them to prevent external payment

⁸⁵ Taira in Dore, *Structural adjustment in Japan: 1970-82*, 115.

⁸⁶ Wade, *Governing the Market*.

problems, to control imports and to direct investment. They gave industries whatever protection they thought was needed through tariffs and import controls and, with the financial aid of the US, started industry growing quickly by ensuring its profitability. Virtually all financing was at first controlled by the authorities, who arranged that industries they wished to help should obtain financing on especially favourable terms. Just as in Japan, the controls over the economy had later to be loosened or given up, though this happened so gradually that the authorities could even in the 1970s determine, when they wanted to, where investment took place and what was imported, by when Taiwan's industry had for some time been growing fast and producing complex goods with up to date knowledge.

The capacity of the education system to train scientists and engineers was expanded, so that by 1971 Taiwan had roughly eight engineers per thousand of population, more than any other low wage country except Singapore, which had ten. Training of workers at secondary and vocational schools was expanded correspondingly. Production of complex goods was not prevented by a lack of trained workers; when foreign firms transmitting proprietary knowledge needed to provide the training for its use, they could find people with the prerequisites.

At the time the main ways for Taiwan's firms to obtain proprietary knowledge from foreign firms were to buy it or to get it as part of subcontracting arrangements. The only alternative, inducing foreign firms to transmit proprietary knowledge in exchange for access to the domestic market, through joint ventures for instance, could work in a few cases, at best; in an economy as poor as Taiwan, with a population of not quite 15 million, demand would in most cases have been too small. But knowledge bought for production for the domestic market, whether by such a joint venture or by a wholly Taiwanese firm, would not necessarily be up to date; the buying firm might judge that the lower cost of knowledge that was old, though still proprietary, yielded more profit, or the selling firm might refuse to sell any that was more advanced. For the benefit of having versions that might not be much more advanced than what generally available knowledge could give, the cost would be that arising from the exclusion, wholly or partly, of imports of more advanced versions of the good in question and the licence fee. Moreover, the knowledge bought would have to be replaced repeatedly by more recent knowledge. One of the few cases that did work, the imposition of local content requirements on the assembly of Japanese television sets for the domestic market, did so because the local content became part of general electronic components industry.⁸⁷

⁸⁷ Wade, 94.

Since foreign firms were unlikely to sell knowledge that would have allowed Taiwan's firms to become competitors, proprietary knowledge was only to be obtained for production of goods that the high wage country firms intended to stop producing or for subcontracting. The former were goods that the producing firms judged would be so little improved by R&D that low wage competition would soon reduce profitability. In such cases the high wage country firms would not only be willing to sell their proprietary knowledge, but would compete to sell, whilst the low wage country firms that started producing those goods first would obtain the same income as the high wage country firms, less the correspondingly lower licence fees. Thus, production of plastics (PVC) began in 1957 and so did synthetic fibres as a joint venture with a US firm. By being among the first low wage country producers of these goods, Taiwan's firms obtained at first the same income as the high wage country firms. Construction of a steel mill owned by the state began in 1970 after several year of study using foreign consultants, though some small private mills had been operating under protection.

But most of the growth of the Taiwan economy came from subcontracting. So much so, that Taiwan came to be described as 'simply a collection of subcontractors for the American market'⁸⁸, though a large part of the subcontracting was for Japanese firms. It began with electrical and electronic goods. Simple electrical goods were already being produced and radios assembled with high levels of protection when, in 1953, a firm first extended its range of products by getting knowledge and training from a Japanese firm. Several joint ventures with Japanese firms to produce electrical goods followed in the next few years and, on the advice of US experts in 1961, the authorities decided to try to build up electrical and electronic industries.

Soon production of electronic and electrical goods and components under subcontracting arrangements had become one of Taiwan's main sources of income, but one that depended on high wage country firms, who both bought the components and provided the knowledge for their production. The authorities did try to establish in the economy the capacity for designing and making electronic circuits that could compare with those of the high wage country firms, but they did not catch up within the period being discussed. Planning for manufacturing semi-conductors began in 1972 and enough workers were trained to set up a model wafer fabrication plant that started in 1976. Knowledge for designing circuits was obtained at first through a technology transfer agreement with a US firm. Later the authorities attracted Chinese-American firms from Silicon Valley and then induced a European firm to enter into a joint venture with some Taiwanese firms to

⁸⁸ Quoted in Bello and Rosenfeld, *Dragons in Distress: Asia's Miracle Economies in Crisis*.

make VLSI circuit chips. Nevertheless, Taiwan's circuits remained several steps behind those of the high wage country firms.

Another attempt to compete with high wage country firms that failed was in the production of motor cars. In the 1970s the authorities decided to establish in Taiwan capacity for producing motor cars using some proprietary knowledge of high wage country assemblers of motor cars and makers of components. Their attempts 'wobbled' and eventually ended with Taiwanese firms becoming subcontractors for components.⁸⁹

A more successful and restrained attempt to compete was that of precision machine tools. Some Taiwanese firms began early to produce machine tools of low quality and low performance and with time improved both design and quality, becoming eventually exporters of precision tools of high quality. At least in part, the progress depended on high wage country firms for the most advanced components, in this case the numerical control components provided by Japanese firms.

Unlike Japan, whose proprietary knowledge at the end of the War was close to that of the firms of the West, Taiwan's authorities had to allow foreign direct investment and to let foreigners hold management positions. Little proprietary knowledge would have been transmitted without joint ventures or some control by foreign partners. Consequently, much of industry, in particular the firms producing complex goods, consisted of joint ventures and subsidiaries of foreign firms.

Independently of the acquisition of knowledge, foreign direct investment was also sought as a source of employment and income. Taiwan started the world's first export processing zone in 1965 and, because of its success, set up several more in the following years. Investment of this sort is not a means of bringing about the transmission of knowledge, but, before subcontracting became important, it generated employment and income at little cost to the Taiwan economy.

Before subcontracting and foreign investment became the major sources of income in the 1960s and 1970s, however, imports and licence fees for proprietary knowledge had to be financed by exports of simple goods and foreign aid. Cotton textiles were one obvious choice of exports at the start, especially as owners of textile firms on the mainland had brought over as much of their machinery as they could, and they became the main exports for some years. They may not in the early 1950s have been quite the low wage goods they were to become as other low wage countries began to produce more of them and the high wage countries increased the protection

⁸⁹ Wade, *Governing the Market.*, 101–3.

of their own cotton textile industries, but their export were nevertheless subsidised, provided the firms met specified targets.⁹⁰ Similarly, with the domestic production of synthetic fibres and the increase in the variety of textile exports in the late 1950s, Taiwan was early enough to avoid some of the protection against textiles made from synthetic fibres. Apart from textiles, Taiwanese firms produced a variety of simple goods for export, many of them of plastic. In addition, the US probably gave more economic aid in relation to the population than what any other economy received at the time and that allowed output to grow faster.

South Korea

Like the new government in Taiwan in 1950s, the government that took power in South Korea in 1961 decided at once that the country should industrialise as a producer of complex goods. Nearly all its industry originated from the Japanese occupation; Korea as a whole had supplied some of Japan's needs for cotton textile, as well as all as its own and Manchuria's, and it had supplied iron and steel products for Japan's manufacturing. Much of the metal and metal working had been in the North and had been demolished during the Korean War. Little had been added to the industries in the South, notably the textile industry and some of the iron and steel industries, but the damage from the two wars had been repaired with US economic aid. South Korea⁹¹ was also short of trained workers, though some came with the return of many Koreans from Japan after the War. Consequently, industrial production was confined to simple goods. Moreover, Korea relied on US aid under PL480 for its supplies of raw cotton, of which the terms prevented exports of cotton textiles.

Korea's authorities, too, regarded foreign firms as the main sources of advanced knowledge for their economy and used the powers they had to induce them to transfer some of their proprietary knowledge to Korean firms. Their approval was needed for practically all industrial investment, all foreign exchange received by residents had to be sold to them, they controlled imports, set tariffs and quotas, they directed bank lending and provided subsidies of various kinds, including loans at low interest rates. They seem to have concluded early that subsidiaries wholly owned by foreign firms were not a good means for obtaining knowledge they wanted. So, they continued to try to attract FDI, but imposed conditions. They required, in principle, that it not compete in the domestic market with Korean firms, especially in the production of consumption goods, and that it be in the form of joint ventures in which the foreign firms did not hold majority shares. But they kept their

⁹⁰ Lin, *Industrialization in Taiwan: 1946 - 1972.*, 108.

⁹¹ From here on South Korea will be referred to as Korea.

ability to deal with all foreign investment case by case and they modified the 'Foreign Capital/Technology Inducement Laws' as they saw fit.

Poor as it was then, Korea had little to attract investment by foreign firms, despite a population bigger than Taiwan's. The domestic market for complex goods was small and that for simple goods consumed in the country was or was likely to be reserved for Korean firms. High wage country firms seem also to have thought there were not enough trained workers and that the initial costs of a joint venture with a Korean firm would have been too high. Hence, the only reason for investing there was the possibility of using low wage labour for simple assembly operations for export. Some US firms started assembly operations for electronic goods, such as radios, but with little or no transfer of advanced knowledge to Korean firms.

As Castley shows in some detail, this changed with the Korea-Japan Normalisation Agreement of June 1965, which started a permanent pattern of cooperation between Japanese and Korean firms.⁹² Apart from providing \$ 800 million of assistance in terms of goods, services, credits and grants over ten years, Japan undertook to "subcontract to Korea its labour intensive export oriented processing industries".⁹³ Japanese officials were explicit that transferring the production of simple goods and components of complex goods to Korea would benefit their economy by allowing Japanese firms to increase their own production of goods requiring more advanced proprietary knowledge, when they would otherwise have been constrained by full employment. Such transfers would be immediately profitable because of Korea's lower wage and, as the officials were aware, would, when enough of the production had been transferred for final goods to be marked as made in Korea, allow Japanese firms to avoid the restrictions on their exports imposed by Western countries. An expression of this view, though never officially adopted, was the Yatsugi Plan for cooperation between the "capital and technology" of Japan and the "labour and land" of Korea in steel, aluminium, oil, petrochemicals, ship building, electronics and plastics.

Japan's and Korea's economic development over the next two decades bore out the judgment of the Japanese authorities. Korea's output of manufactures grew rapidly, some of it being exports of simple goods, primarily textiles, some of it being the transfer to Korea of the production of complex goods, much of which consisted of subcontracting for Japanese firms and generated correspondingly more income. Most of the knowledge and training needed for the subcontracted production was provided by Japanese firms. Many subcontractors were entirely Korean firms that established long lasting partnerships with Japanese firms, and many were

⁹² Castley, *Korea's Economic Miracle*.

⁹³ Quoted by Castley (p.87) from K.B.Kim.

joint ventures of Korean and Japanese firms. 'Most foreign direct investment came from Japan. During the whole period 1962-79, there was a total of 617 cases of direct investment from Japan, followed by the US with 97.'⁹⁴ Leaving Japan aside, 'Korea has not relied upon direct foreign investment as a source of foreign technologies ... in the main Korea has purchased both the capital goods and the know-how ...'.⁹⁵

Manufacturing of this sort grows by progressing from the simpler to the more complex, which seems to be what Japan's officials anticipated. Korea was no longer prevented from exporting cotton textiles by PL480 and began to export to Japan, the US and other economies. The output of Korean textile firms grew, but Japanese firms also invested in textile production in Korea. By the early 1970s over two thirds of Korea's textile exports went to Japan. There was an orderly transfer of production as Korea's textile producers progressed to more complex products, notably synthetics and blends, and the Japanese textile industry abandoned step by step less complex products and specialised in the most complex ones.

As happened in Taiwan, textiles provided income while the production and export of complex goods were still small. Korea's authorities seem to have regarded them in this way, as a stopgap; they tolerated a reluctance to adopt new production techniques that required training of workers or to diversify output, even on the part of some of the bigger textile producers,⁹⁶ that they did not tolerate in industries they considered would generate income growth in the long run.

Korea's steel mill is another case of Japanese firms transferring production, presumably because the firms concerned judged that a successful venture would let them produce less of some items, which would have been costlier to import from the Western countries, and more of some others. After a consortium of the US and European countries had declared a steel mill not to be economically feasible, a group of Japanese firms agreed in December 1969 to help Korea build one and provided the financing, proprietary knowledge and training. Production started in 1973 and the mill was profitable from the start, though the capital costs, infrastructure and services, such as ports and railways, were subsidised by the state. It was expanded repeatedly in the next few years.

Textiles and steel aside, most instances of such transfers were subcontracting arrangements between, on the one side, purely Korean firms or joint ventures of Korean and Japanese firms, and, on the other, Japanese

⁹⁴ Castley, *Korea's Economic Miracle.*, 154.

⁹⁵ Enos and Park, *The Adoption and Diffusion of Imported Technology.*, 39.

⁹⁶ Amsden, *Asia's Next Giant.*

firms, and they covered, among other things, electrical and electronic goods, chemicals, plastics and machine parts. They were usually long lasting partnerships in which the Japanese firms provided proprietary knowledge and training and which mostly progressed to more complex activities, though the knowledge and training did not go beyond the purpose at hand to allow production that the providing firm wished to keep for itself. In the early stages the Japanese partners sometimes had to help Korean firms, especially some that were not big, to acquire generally available knowledge that these firms had not had the time to acquire and to train their workers accordingly.⁹⁷

But Korea's authorities and its conglomerates, the chaebols, did not confine themselves to subcontracting and cooperative arrangements; they differed from Taiwan's firms in that they also began to produce complex final goods, where necessary with brand names, that competed with the production of high wage country firms. At first the goods were cheap and technically backward. For example, in the 1970s some East Asian countries were markets for audio systems made by Korean firms using knowledge that they acquired on their own and the US for cheap microwave ovens also made by Korean firms.⁹⁸ These conglomerates were bigger and more diverse than any in Taiwan and had been fostered by the authorities because they made it easier to direct activities, especially to start production of new types of goods or to meet performance targets. Their size allowed them to establish some R&D capacity and buy the proprietary knowledge of high wage country firms, whilst their diversity let them both spread their risk and cross-subsidise their new activities.

A Korean motor car industry was decided on by the authorities already in 1962, but the first attempts failed. It was in 1973, as part of a bigger plan, the "Heavy and Chemical Industrialization Project", that the authorities were able to direct the chaebols, which were bigger by then, according to better prepared plans.⁹⁹ Even then they almost failed. Success depended on obtaining adequate proprietary knowledge, whereas the venture that the authorities favoured was a joint venture with a high wage country firm that was also the sole supplier of proprietary knowledge. That which succeeded best was one they did not at first favour, but which was independent of foreign firms and bought proprietary knowledge from several different sources. The performance of the first improved once the chaebol involved took complete control¹⁰⁰, though it was wholly taken over by the same foreign firm as a result

⁹⁷ Cyhn describes several cases and brings out the mechanics.

⁹⁸ Cyhn, *Technology Transfer and International Production*, 250.

⁹⁹ Doner et al., "Industrial Competitiveness of The Auto Parts Industries In Four Large Asian Countries."

¹⁰⁰ Kim, "The Dynamics of Technology Development: Lessons from the Korean Economy", 148.

of the crisis of 1997 and fared badly after. A third also failed in 1997 and had to be taken over by the chaebol that owned the most successful venture.

There seems to be no account explaining why the few high wage country firms that could supply the proprietary knowledge supplied it to a firm that eventually became a competitor. One conjecture is that they did not believe that the chaebol would succeed. Production in low wage countries of motor car components by subsidiaries of high wage country motor car producers and by subcontractors was only just beginning, so that the only relevant experience of Western firms was that of the assembly plants they had built in several low wage countries, none of which had gone beyond assembling components supplied by the firms with, at most, minor local content. Since the knowledge was being provided by several firms, none may have been aware of how much was being transmitted. It is also possible that the knowledge was costly and not up to date, but adequate for a start since the chaebol at first produced for a protected domestic market and was directly or indirectly subsidised. When exports started in the late 1970s, they may have been possible because, even if the models of the cars were technically less advanced than those of high wage country producers, the low wages allowed them to be cheap.

In other cases chaebols obtained proprietary knowledge from Western firms that were closing or in financial difficulties and, therefore, willing to sell at prices that would allow the users to be competitive. When a Korean firm tried to start building large ships and could not get proprietary knowledge from Japanese shipyards, it obtained proprietary knowledge, in particular blueprints, from Western shipyards, which were closing because of Japan's competition. Its first attempt failed, but it soon produced hulls that could be sold to foreign shipping companies.¹⁰¹ In the 1970s a chaebol starting to design and produce integrated electronic circuits obtained the proprietary knowledge of US firms that needed financing to continue and were willing to sell proprietary knowledge or let themselves be taken over. Later, Korean firms took over or set up R&D capacity in high wage countries to be able to employ scientists and engineers from there.¹⁰²

Knowledge acquired like this would not have enabled the Korean firms to become competitors with their own brand names had they not also had their own capacity for R&D. Even if what it acquired was up to date at the time, for the firm to establish a brand name and compete outside the country it had to generate its own knowledge in step with the established producers. Only big conglomerates and state R&D institutions had the means of doing that at the time. This does not mean that the chaebols generated on their

¹⁰¹ Amsden, *Asia's Next Giant*.

¹⁰² Kim, "The Dynamics of Technology Development: Lessons from the Korean Economy".

own proprietary knowledge as advanced as that of the high wage country firms with which they competed. It means that, once they had the capacity to generate parts of the knowledge they needed, they were better able to obtain the rest from high wage country firms. Once Korean firms showed they were going to be permanent competitors and, if need be, to export less advanced products at low prices, the high wage country firms, especially Japanese firms, sold them much of the proprietary knowledge they wanted.

Thus many components of their motor cars were supplied by Japanese firms, as were the engines and some of the machinery of their ships. A chaebol produced more flat panel displays than any other firm in the world, but some of the most advanced electronics came from Japan. High wage country firms could also agree to share R&D, as has happened throughout the motor car industry. In some cases having the capacity to do the R&D was, itself, enough to reduce the prices of imports. When a state run R&D institute succeeded in making fibre optic cables, the main supplier reduced its prices. Much the same occurred with some chemicals.¹⁰³

Royalty payments by Korean to foreign firms show how much Korea's output of complex goods increased. They are a comprehensive, if rough, guide, for subcontractors and joint ventures had to pay for the proprietary knowledge transmitted by the partner firms of high wage countries, though usually at lower rates than wholly Korean firms that bought proprietary knowledge for which they did not subcontract. Since they are normally proportional to the output of the goods concerned and cease when the knowledge is no longer used, they combine the technical progress and the resulting output. Castley¹⁰⁴ quotes data originating with the Korean Economic Planning Board and gives the payments as \$ 1.3 million in 1968 and \$ 19.0 million in 1974. According to Bello and Rosenfeld¹⁰⁵, licensing and royalty payments rose from \$ 58 million in 1977 to \$ 1.2 billion in 1989.

China

A description for present purposes of how China industrialised can be confined to how Chinese firms acquired advanced technical knowledge from the firms of high wage countries and can be brief. China is so far the last of the East Asian countries to have industrialised by this method and resembles Taiwan in some respects; there were no benevolent countries to help it and yet it had to find ways of inducing foreign firms to impart proprietary knowledge to its firms.

¹⁰³ Kim, 159.

¹⁰⁴ Castley, *Korea's Economic Miracle.*, Table 4.5.

¹⁰⁵ Bello and Rosenfeld, *Dragons in Distress: Asia's Miracle Economies in Crisis*, 114.

Its political leaders who began to open the economy to foreign trade and investment around 1980 knew that China's technical knowledge was behind that of the high wage countries and also understood that, if China's firms did not acquire the proprietary technical knowledge of the high wage country firms, the supply of complex goods would come from imports and local production by foreign owned firms. After it took power on the mainland in 1949 the communist government had had to rely on the technology it could obtain from the Soviet Union several years before and its own R&D. But it invested more compared more of its income in education at all levels than other low wage countries, though episodes like the "Great Leap Forward" and the "Cultural Revolution" seem to have been in part attempts to escape or even deny the importance of advanced technical knowledge and to replace it with the fervour of the masses.

China's advantage, as the political leaders understood, was size; it was the economy with the biggest population in the world. They also saw that the economy had the advantages of low wages, virtually full literacy, ample numbers of scientists and engineers and good infrastructure. For foreign firms the prospect of a large market that had had little of the conventional consumer goods and the possibilities of manufacturing for export with low wage labour were attractions. But the Chinese authorities kept control over imports, exports and investment and, when they could or thought it appropriate they saw to it that a foreign firm that wanted to sell to the Chinese market was obliged to produce locally in a joint venture or some sort of partnership with a Chinese firm, to which it had to impart proprietary technical knowledge. They also prevented foreign firms from taking control of Chinese firms, just as Japan and Korea had done before.

The intentness of China's leaders on their economy's firms acquiring technical knowledge was evident from the criticisms in the high wage countries that Chinese firms violated patents and copyrights and that they stole intellectual property with the connivance of the authorities. High wage country firms also objected to "... being forced to transfer business and technical know-how to Chinese companies in exchange for market access."¹⁰⁶ It was a bargain they nevertheless accepted; the market was too promising and the competition for access too great for them not to join in, as shown by the example of the German car makers, who were among the objectors but for whom China has become one of the biggest markets.

An early indication of the success of these methods was the extent to which high wage countries firms directed their foreign direct investment to China during the three decades after China's opening began. Because of that investment a great part of China's exports quickly came to consist of goods

¹⁰⁶ "German Industrialists Lambast China." Financial Times. April 19, 2010. (p.1)

that the firms of high wage countries produced for their own use and trade in which was not as restricted as the textile and garment exports of other low wage countries, though even China's textile and garment exports could grow faster than those of other countries.

But the main success was the rate at which the variety of complex goods that China could make increased and, with that, the extent to which Chinese firms could produce under their own brand names for export. There is no need to belabour the point now that China has become the second biggest economy in the world.

China's success may be hard for other low wage countries to emulate, quite apart from the question of size and the degree that the economy's past, especially its training of scientists and engineers and its industries, however backward, prepared it for later success. China had no foreign debt, unlike most low wage countries, which depend on foreign financing, often merely to service their debt, and so did not have to follow the advice of creditor multilateral and bilateral institutions. Its exports were also helped by the credit expansion in the high wage countries as a whole, especially the US through the late 1990s until the economic crisis of 2008. The combination of a high saving rate in China, probably partly the consequence of fast growth, and low saving rates in the US and some other countries allowed fast export growth and trade surpluses. As yet the institutions providing external financing to the low wage countries have shown no sign of advising these countries to follow the kinds of policies that the four economies just discussed have followed with success.

Something these four economies have in common in addition to having become technically advanced is radical social change after World War II. Japan had a drastic land reform while occupied by the US. Japan's departure from South Korea removed a long-standing structure of government and the ownership of a large part of the country's wealth. China had its revolution in 1949. Taiwan was invaded by Chinese fleeing the revolution and the old social structure, which had grown under long Japanese occupation, gave way to a new one. These were all social revolutions and, except for China, some of their success can be attributed to the wisdom of the American economic advisers who initiated or guided some of the social change and early economic policies, people imbued with the ideas of the New Deal. Their purpose was to protect these economies against communism by getting them to prosper through suitable economic policies, financial aid and trade opportunities, and it is doubtful that they would have been as successful if they had applied the doctrines of the later generations of advisers and economists.

India

India had no social revolution on becoming independent in 1947. Nonetheless, before Taiwan or Korea, its authorities decided that the country should have an industrial economy. It had a less diverse industrial sector than had been left by Japan in Taiwan or Korea, no production of complex goods and, therefore, no firms with technical knowledge that was not generally available. Its population was several times that of Japan and at least as poor, the literacy rate was low, there were few engineers and the education system could train only a few scientists or engineers.

Industrialisation as conceived of by the authorities included the capacity to produce complex goods, capital equipment in particular. India was too big and its political leaders too ambitious to be content with exporting cotton textiles and other simple goods, however big and modern its textile industry may have been. It was, in any case, unlikely that such exports could grow fast enough over decades to provide for the imports the capital and other complex goods needed by a country of its size, especially with the prevalence of trade restrictions at the time. What seemed then to be the alternative was Russia's industrialisation as an isolated economy in the 1930s. Feldman, who provided the reasoning of Russia's plans, argued that creating greater capacity for producing capital goods now would allow more consumption in the future.¹⁰⁷ His reasoning was elaborated on by several Indian economists at various times, among them P. C. Mahalanobis, G. Mathur, K. N. Raj and A. K. Sen, who accepted the premise that international trade could only be an adjunct to the economy, though it did not result in as extreme a concentration on capital goods production in India as in Russia.

For several decades India's economic growth was slow compared to the rates reached by the other countries discussed above. One reason was that the country had a system of administration and social structures that were hierarchical, stable, and designed for rule by an outside power. As usual with empires, the British had kept social change to a minimum, especially after the uprising of 1857. There were ethnic groups and castes engaged in commerce and manufacturing, but the education system, civil service, judicial system and the segmentation of economic activities by ethnic group, caste and class were ill suited to industrialisation. Since the course chosen involved extensive regulation of investment, foreign trade, commerce and financing by a hierarchical civil service set up to administer a large country conservatively, the regulation was bound to become inefficient and corrupt. Another reason for slow growth was that those who made the decisions did not try to acquire for the economy the latest techniques of the high wage country firms, but wished to create industries with the production techniques

¹⁰⁷ Dobb, *Soviet Economic Development Since 1917*. 6th Ed., 332 & 360–61.

Indian firms and state agencies could develop themselves. As explained above, the Indian economy could be viable, but was confined to production techniques and complex goods that had to be protected against foreign competition because they were out of date and produced few exports that were not simple goods.

Detailed foreign exchange controls were necessary, both as part of the protection of domestic production and for directing investment by firms to the production of capital or other complex goods. At the time India became independent foreign exchange controls were normal around the world; almost all countries, apart from the US, had foreign exchange controls and even the US had trade barriers. But, whilst the high wage countries gradually removed these controls, India had to keep them. Also normal was the belief that infant industries had to be protected against foreign competition in their early years. Most Indian economists and government officials accepted the argument that a firm starting production of a complex good needed time to become efficient enough to compete with established producers and that, without protection, the likelihood that it would fail would deter investment in it.

At the same time the authorities wanted Indian firms to be able to design complex goods and the production processes involved, not just repeat versions they had been making. In practice this meant, first, that Indian firms used generally available knowledge, augmented perhaps by R&D from within the country, but did not obtain advanced proprietary knowledge from high wage country firms, and, second, that complex goods were only imported if they could not reasonably be produced within the country or if there was need for there to be some competition. FDI was allowed when it seemed that Indian firms would not be able to produce the same goods and was not, therefore, used as a means of obtaining proprietary knowledge. High wage country firms were not often attracted to India; they often felt that the market was small because of the poverty, the risk of nationalisation and administrative interference were too great and that they would be discriminated against, especially if Indian firms competed against them.

Income from manufacturing rose as the numbers of people employed rose and as the goods and production processes improved, though its industries were out of date compared to the high wage country firms. But the growth was too slow to yield full employment and the real wage stayed low. India continued to export simple manufactures, which, along with foreign aid, allowed it to avoid scarcities of imported raw materials, capital equipment and consumer necessities that could have prevented industry from functioning.

For a long time India's authorities seem not to have been aware, as were those of Taiwan and Korea and, later, those of China, that they had to find

ways of inducing high wage country firms to transmit proprietary knowledge to Indian firms. They could, as did China, have exploited the size of the country by making foreign firms' access to its markets conditional on joint ventures with domestic firms or on meeting specific local content requirements, with the associated provision of training and knowledge to Indian suppliers. They could also have fostered subcontracting arrangements by adapting procedures for foreign exchange, import/export, customs and taxes. Instead, when Indian firms wished to buy proprietary knowledge from high wage country firms they had to obtain the approval of the authorities and could not be sure that the foreign exchange for paying the licence fees would be available. Otherwise firms had to rely on generally available knowledge, for investment in R&D by Indian firms was small.

The results were not failure, though economic growth was slower than need have been and much human effort was wasted because of outdated methods producing outdated goods. Indian industry and the income it generated did grow, whilst the numbers of workers with training at various levels, even some at the levels of the best Western universities and research institutes, increased relative to the population and to employment. This allowed a form of subcontracting that was not handicapped by the complications of subcontracting goods production, namely the provision of services by people. High wage country firms wanting to transfer production processes (taken here to include after sales services) began to give tasks that could be performed just with transmitted data to low wage country firms. Indian firms began to process insurance claims, do specific kinds of medical diagnosis and help customers of high wage country hardware and software suppliers solve technical problems. It helped that transmitting the data for some of these services, which could, in principle, have been done by sending documents back and forth by airmail, could, by the time that firms began to envisage such subcontracting, be done virtually instantaneously. It also helped that the Indians with the appropriate training could speak English, the language of the US, the main source of the business at the time. Other services provided in India did not use such transmission, for instance medical care, ranging from surgical operations and hospital care to diet and exercise centres, could be cheap and good enough to attract people from high wage countries.

Apart from services, the accumulation of scientists and engineers and the growth of industrial production allowed Indian firms to export less advanced versions of complex goods, a well known example being generic medicines. Several firms had learnt how to make medicines whose patents had expired or were soon to do so, or in the case of the treatment of HIV, to sell them with the agreement of the patent holders.

India began to open its economy in the 1990s. It had to begin reducing trade barriers if it wanted to keep normal trading relations with the high wage countries, though the barriers remained higher than those of most low wage countries. Moreover, the desire for freer movement of capital, fewer restrictions on trade and less administrative control in general had grown. More economists had learned orthodox theory and institutions like the World Bank and the IMF had propagated its tenets among government officials. Besides, economic growth had been persistently slow and many thought that change gave better prospects.

Greater openness has had the effect of letting Indian firms acquire more up to date knowledge and that has made the economy grow faster. Indian firms became better able to acquire proprietary knowledge from foreign firms as changes in the regulations made collaboration in the form of joint ventures and transfers of technical knowledge easier and, for foreign firms, India had become a big market. As already mentioned, on occasion Indian firms obtained advanced proprietary knowledge by taking over high wage country firms. Several decades of industrial growth have allowed some firms, of which a few originated in the nineteenth and early twentieth centuries, to become conglomerates that have the size and capital to obtain foreign financing when they want. India's exports are not composed of complex goods in proportions as high as those of the East Asian economies discussed here nor has the rate of growth of its exports been as high, but it is to be expected that, barring greater obstacles to international trade, the component of complex goods will increase.

CHAPTER 3

THE BALANCE OF TRADE AND THE EXCHANGE RATE

1. NATIONAL PRODUCT, EXPENDITURE AND THE TRADE BALANCE

The Balance of Trade

Preliminaries

Up to now the discussion has been confined to the pattern of trade and production and to how it is determined either without reference to the balance of payments or with the balance of trade as part of the mechanism that results in the pattern. By their assumptions that lead to single period or intertemporal equilibrium with expectations realised, the models and theories discussed so far preclude any discussion of problems of the balance of payments within their confines. Since such problems do occur, the causes must be sought outside these confines. From them would follow the means for influencing a country's balance of payments should there be reason to do so. Any explanation ought to begin by specifying the theory or model of trade and proceed by specifying how the assumptions have to be modified, but this is not the procedure textbooks and theoretical papers normally follow; usually they use one set of assumptions for the determinants of trade and production and another set for the balance of trade, without explaining how the two sets of assumptions are related or are to be reconciled. Most textbooks are divided into two parts along these lines.

In what follows the procedure is to discuss the balance of payments by modifying the assumptions necessary for intertemporal equilibrium and proceeding step by step using the conventional categories of economic accounts. The discussion is conducted in terms of the country's residents, namely the households and firms meeting some residence criteria, which need not be specified here, and the government. All that is needed is that the criteria require the individual or firm to be in the country for a period comparable to the periods in terms of which the discussion is carried out, a year or half a year, and that they be applied uniformly to all countries, so that

all individuals and firms are resident in some country and none are resident in more than one country at any time.

All the production of goods and non-factor services by residents of a country in that country constitutes the country's gross domestic product (GDP).¹⁰⁸ Its market value over any period is equal to the wages paid by and profit of resident firms and government plus income from self-employment of residents plus indirect taxes on production. Deduction of the profit and wages accruing to non-residents from their ownership of or employment by firms resident in the country gives the total income from production of all residents. The pay of individuals resident in one country who work abroad without changing residence is considered factor income of their country of residence and is a cost of production in the country where the work takes place. Ownership of firms can be the ownership of some shares in the firm or total ownership and the income to non-residents from it is also factor income for their countries.

Income that individuals receive from employment, including self-employment, and from productive wealth in the form of ownership or shares of firms or of land, buildings and other productive assets can be termed their income from production. Individuals can have non-production income from loans and debt bearing interest and thus have net income from debt service, i.e. receipts less payments of interest and principal on debt outstanding, as well as from buying or selling productive assets and receiving or making transfers. Each firm obtains income from production after paying its employees, and that is profit gross of depreciation, to which can be added profit the firm receives from productive assets it owns separate from its main activity. Individuals and firms pay direct taxes on income and wealth. The government also receives income; it consists of the direct taxes paid by individuals and firms, indirect taxes on domestic production and imports, transaction taxes on sales other than of new goods, fees for services the government provides and income from assets the government owns, including shares in the ownership of productive capacity. Contributions of employers, employees and government for benefits of employees, such as medical and unemployment insurance and pensions, are sometimes classified as taxes, but are really part of the pay of employees in the form of entitlements to eventual transfers. For simplicity these contributions and transfers will be left out of what follows, and so will forms of wealth that have market value but may be considered unproductive, such as antiques, art objects, original manuscripts, stamp collections and so on.

From here on references to the income of individuals or firms will be to income after direct taxes and the profit of firms is taken to be gross of

¹⁰⁸ For brevity the term goods is used from here on to include non-factor services.

depreciation, unless otherwise specified. Since people are employed as individuals but live in households, often with dependents, their expenditure can be ascribed to households. The difference between a household's expenditure and its income from production is the sum of net debt service, new borrowing or lending and sales or purchases of assets constituting wealth.

By definition, all imports are imports of residents and the expenditure on goods by non-residents, whether abroad or in the country, counts as exports. Hence, the amount by which the income of the residents and non-residents of a country from production in a country exceeds the expenditure of residents is that country's trade balance; it is the excess, positive or negative, of the market value of the gross domestic product over the value of goods and services on which the expenditure was made.

Several complications are better left to manuals on national accounts. Among them are special arrangements for foreign firms, such as export processing zones and off-shore status, military bases of foreign countries and so on. In practice some individuals do not meet the criteria for residence; some are resident in more than one place, often because countries are not entirely consistent about their criteria, and some individuals avoid meeting the residency criteria of any country. The latter include refugees, illegal migrants and wealthy people avoiding taxes by not staying long in one place. It is practically impossible to observe directly the daily consumption in a country of non-resident individuals, tourists in particular. In the special case when non-residents remit large amounts home, as with migrant labour in the Middle East, the remittances are taken as their factor income. In principle, national and balance of payments accounting is adjusted in various ways to allow for these cases with consistency between countries and they do not alter the argument that follows.

Expenditure can be divided into consumption and investment (gross fixed capital formation (GFCF) and increases in stocks)¹⁰⁹, which can be divided further into private sector consumption, private sector investment, public sector consumption, which is considered the same as government consumption for present purposes, and investment by the government. In accordance with common practice, investment by public sector firms is included in private sector investment and government investment is part of the government's budget.

¹⁰⁹ Education is also investment in the economic sense, but is normally treated as consumption in national accounts. A reclassification of education and expenditure on it as investment would not affect the relation between the trade balance and expenditure. The same is true for health, though it is less directly important here. The effects of education on future production possibilities and costs will be discussed separately.

What a resident spends on goods in a period is the sum of that resident's income after income tax and net borrowing in that period, less net acquisition of wealth assets. Debt and debt service and sales of assets between residents of a country net out, so, according to whether, in the aggregate, net borrowing and net sales of assets in that country add to less or more than net payment on debt service, the residents spend less than or more than their incomes from production and that country has a surplus or deficit on its trade balance.

For convenience the discussion is conducted as though time were a sequence of periods. In any period the various combinations of goods a country can produce, its production capabilities, are assumed to be given and to be limited by the capacities of its stocks of capital equipment, which do not change during the period. The numbers of workers firms employ when producing to capacity are given by the types of productive equipment the firms use and, when production is not at capacity, some of the equipment is not used to capacity, is idle, and some of the workers are unemployed or underemployed. With the additional assumption, that every good is produced competitively in several countries, the discussion here is consistent with the model of Chapter 1.

The balance of trade: expenditure and competitiveness

Each period the residents of a country plan their spending and firms plan their production for that period. But there is no assurance that these plans are mutually consistent in the aggregate. Just as the amounts the residents of any country intend to spend in a period do not, in the aggregate, have to be equal to the value of production that the firms in the country plan for the period, there is no *a priori* reason that the total that the residents of all countries intend to spend should be equal to the total value of production planned by all the firms. For example, the residents of some countries may plan to spend more than their countries' production capabilities allow at the going prices and those of others to spend less, whilst the firms of all countries plan to produce to the limits of their capacities. It is possible that the trade deficits implied by the former countries' plans and the surpluses implied by the plans of the latter countries match, so that all plans are realised. Then the plans were mutually consistent. But it is also possible that the plans are not mutually consistent and that the deficits implied by the spending and production plans of the former countries add up to more than the implied surpluses of the latter countries, or add up to less.

One question, therefore, is how are such inconsistencies within a period resolved? They are resolved within the period, since the trade balances of all countries add to zero, and, when they are resolved, some expectations turn out to have been mistaken and some plans are not realised. Then, the possibility has to be allowed for that production in some or all countries is

not at the limits of the capabilities. Firms may not produce to capacity, though they planned to, and it is also possible that, knowing this can happen, firms at times do not plan to produce to capacity. This in turn means that the question of capacity utilisation is separate from the question of the mutual consistency of the aggregate spending plans of countries' residents and the production plans of countries' firms. For, it is still possible for the firms of all countries, taken together, to plan to produce below capacity and for their production plans and the spending plans of the residents to be mutually consistent, so that the expectations of residents about their incomes from production and the expectations of firms about spending turn out to have been correct. It is taken here that individuals, firms and the authorities who try to regulate the functioning of their countries' economies want, as opposed to expect, production in their countries to be at the limits of the capabilities.

The authorities can also be assumed to want to prevent a second type of inconsistency, one that occurs between different periods, namely when a country's residents have incurred debt to foreigners denominated in foreign currencies and cannot obtain enough of these currencies to service it. In other words the country has a balance of payments crisis. Here the obligations incurred in some periods turn out to have been inconsistent with the later trade balances and ability to borrow or to sell assets.

To keep to the first kind of inconsistency and how it is resolved within the period, the elements that have to be considered are what the residents of each country expect their income from production to be, what they intend to spend, the spending that firms expect and the plans they make for the period, and the resolution occurs through divergence of outcomes from intentions and expectations. All these elements being interdependent, it is convenient to fix one of them to avoid the descriptions of the mechanisms by which inconsistencies are resolved from becoming too involved. So, it is assumed that the spending plans of all residents are made period by period in nominal terms and are always carried out.

In the case just mentioned, when the amounts by which the spending plans of the residents of some countries exceed the value of the production planned by the firms of those countries, i.e. the implied trade deficits, add up to more than the amounts by which the value of production planned by the firms of the other countries exceed the spending plans of the residents of those countries, the implied surpluses, the deficits of the former countries will be smaller than was implied, some perhaps having surpluses instead, and the surpluses of the latter will be greater than was implied, for the surpluses and deficits must be equal in the outcome. Since it is the value of production that is assumed to adjust, one part of the mechanism is that the value of production of the former countries is greater than the firms, in the

aggregate, had planned, and the other part is that the value of production of the other countries is less. To the extent that their production capacity prevents firms from increasing the value of their output by increasing the quantity, the prices rise. In other words, the two mechanisms that resolve the inconsistency are price rises and changes of quantity from what was planned, the latter being constrained by the production capabilities of the countries, and a combination of both is also possible.

If the implied trade surpluses are greater than the implied deficits, aggregate demand is less than the value of production the firms of the various countries plan. Some countries that have implied surpluses will have smaller surpluses than were implied, or even deficits, and the deficits of some of the others will be bigger than were implied. In either set of countries the total value of production is less in the outcome than the firms had planned. In the aggregate, production adjusts to demand. Some firms in some countries may produce as they planned and the possibility that some firms produce more than they had planned cannot be excluded *a priori*, but they are not enough to offset the others. In the countries where the value of production in the outcome is less than the firms had planned income from production is less than what the firms had anticipated and residents' expectations, to the extent that they accorded with the firms' plans, turn out to have been wrong.

In this case the inconsistency can be resolved by lower prices; the value of production is correspondingly lower and in some cases that may suffice for it to equal expenditure, in which case the trade balances add to zero. Prices need not be uniformly lower; some prices can fall more than others, the prices of some goods may not change and it cannot be excluded *a priori* that some prices even rise. But they fall in the aggregate and the income from production of the goods of which prices are lower is less than the firms had planned. Alternatively, the quantities that firms produce can be lower. Here the lower limit, corresponding to the upper limit of capacity constraints, is no production.

No distinction has been made between countries, so nothing can be said about how particular countries fare. In some countries production may be much below what their firms had planned and in others it may be the same as the firms had planned or only a little less. At this point the possibility cannot be excluded that countries fare differently at different times; that a country whose production fell little on one occasion falls a lot on another, whilst the reverse is true for another country. Alternatively, it may always be the same countries of which production falls most on each occasion and others the ones of which production always falls least. In this latter case the differences have their causes in the characteristics of each country's economy, usually subsumed under the term "competitiveness". A third

alternative is a combination of the two, some countries are permanently more or less competitive and the consequences of inconsistencies on others depend on the specific circumstances, which is to say that the relative competitiveness of countries can depend on the circumstances.

Given the exchange rates and nominal wages of all countries, the competitiveness of any one country is the competitiveness of its firms in the aggregate. One country can have several firms that are especially competitive in one type of good, say machine tools, and another in, say, shipbuilding. Then, when a reduction of expenditure creates an inconsistency of the kind being discussed, and reduces demand for machine tools first, it may leave the former country unaffected and cause a fall of the output of machine tools in the latter. But, should a further reduction of expenditure especially affect shipbuilding, the former country's shipbuilders will be more affected than the latter's. Countries often have a number of firms that are particularly competitive in the production of some goods in this way. Some examples that have been given are German cameras in the 1950s and Japanese cameras after that, Swiss watches and French perfume. Other examples are the electronic industries of China, Japan and Korea, the aircraft and films of the US and Scotch whiskey. Other countries may have individual firms that are among the most competitive in the world, each producing a different good. Thus a country can have one firm producing shoes, another home stereo equipment and another toys, each competitive enough to be among the least affected by a reduction of expenditure that reduces demand for the goods of the kind it produces. Differences of wage rates can be partly offset by subcontracting with firms in low wage countries, the design belonging to the firm holding the brand name.

It is assumed here that the competitiveness of countries does not change in the short run, which is the main concern when discussing the balance of payment. This allows the competitiveness of countries to vary according to the circumstances, the third alternative. There may, for instance, be a fall of investment without a change of consumption on one occasion and a fall of consumption on another. To this extent the consequences of inconsistencies in the spending and production plans of the different countries' residents depend on the details of the inconsistencies. As long as the discussion of the trade balance is confined to the short term, in the sense that the circumstances can be taken as given for a change of expenditure or of exchange rates at the time, this does not matter. It matters in discussing the longer run and repeated balance of trade surpluses or deficits and, in this respect, the evolution of trade since 1945 gives reason to believe that some countries are more competitive than others.

Inconsistencies of the first kind can, therefore, result in countries not having the trade balances the authorities wanted and some countries in not being at the limits of their production possibilities. Hence, the two questions, is it possible that all countries have the trade balances their authorities wanted and is it possible to combine this with all countries being at their production possibilities limits? A third, more complex question is, if some countries' authorities are willing to reach an agreement on expenditure and trade balances and to abide by it, do they have any means of inducing the authorities of the other countries to do likewise?

Competitiveness and coordination

In answer to the first question, inconsistencies of the first kind can be avoided if, in each country, the authorities have the powers to determine the aggregate expenditure of the residents. Up to now each country's total expenditure was taken as given. Assuming that the country's authorities can determine it is a way of separating the primary concern here, the consequences of the expenditures of the different countries, from the secondary concerns of how and to what extent they can be determined by the authorities. Circumstances under which the authorities are unable to bring about the total expenditure they want do occur and sometimes the authorities do not take the requisite measures for theoretical or institutional reasons, but these questions of macroeconomic demand management are mostly beyond the scope of this work and are only referred to when necessary in the following.

Assuming the authorities have these powers, they can in each country have the trade balance they want if they and the authorities of the other countries can forecast the value of production in their countries, agree on a mutually consistent set of trade balances and coordinate expenditure accordingly. Coordination like this can also prevent the second type of inconsistency by taking account of debt service obligations in foreign currency, prospective capital movements and the international reserves of each country. For instance, some countries with debt service obligations in foreign currency may need to have trade surpluses to avoid balance of payments crises. Such a country may not be able to borrow enough to cover its obligations, or its authorities may judge that more foreign debt will cause a worse recurrence of the situation. In forecasting the inflows and outflows of factor payments to workers and for capital and in taking into account miscellaneous expenditures that all countries can have, such as royalties, patent and franchise fees, dues to international organisations, diplomatic representation, etc., they may conclude that the net capital flows and the country's international reserves will not suffice. Foreign investment in the country may not offset the outflow of foreign capital and, depending on how

well the authorities control the capital account, domestic capital can also leave. If there is such coordination, some other countries plan their expenditures for the country in question to have the required trade surpluses.

The answer to the second question is conditional; the unused production capacity must not be redundant, meaning that the demand for the specific goods it can be used to produce can never be enough to use it fully. With this condition, if a country is not at the limits of its production possibilities the reason is that it has some production capacity that is not used because the prices of the goods it can produce are too low, given the costs. If the goods in question are tradable and the costs of production are lowered relative to prices, which can be done by reducing nominal wages or devaluing, the country's output of those goods increases, provided the competition in other countries does not offset the cost reduction. If the trade balances and spending by the residents of each country have been agreed, any increase of production in a country following devaluation or wage reductions has to be matched by expenditure in that country that keeps to the agreed trade balances. But if, to take the simplest case, the unused capacity was confined to one country and could only be used for producing one good, part of any greater output is presumably exported and the increased expenditure goes to the import of a variety of goods. Production of that good in some countries is reduced. It can be assumed that the production capacity for that good that falls idle can no more be put to other use than the production capacity that was idle originally. In that case the total capacity for producing that good is greater than can be used with the agreed trade balances. Perhaps a different agreement on spending and trade balances by which countries with greater propensities to buy the good in question spend more would lead to greater use of this capacity, but it is possible that no plausible agreement can result in the full use of that capacity, some of which is, then, redundant. If the good in question is not tradable, lower nominal wages or devaluation are unlikely to cause its output to be higher unless spending is increased. But, just as with the case of tradables, increased spending can be expected to be spread over other goods as well and to worsen the trade balance.

Redundant capacity of this sort occurs often. Cotton textiles has already been given as an example. Another good of which production capacity may have exceeded likely demand for some time is the motor car. In these cases the redundant capacity was created by firms independently trying to increase their production. But capacity can also become redundant because consumer preferences change or a new good displaces an old one. Men no longer wear hats as they used to until the late 1950s, film is no longer used for ordinary photography and quartz watches have turned clockwork watches into expensive jewellery. These changes resulted temporarily in

unusable production capacity. Redundancy of this sort can affect the trade balance, but whether or not a country adapts to the changes is a question concerning the pattern of trade and the long term. Here firms are assumed to judge the capacity they need for the current and next few periods well and when their production is below capacity it is because aggregate expenditure has been cut back, though the composition of production would otherwise have matched demand.

Judging by the lack of coordination, the answer to the third question seems to be that there are no means by which some countries' authorities can induce the authorities of other countries to coordinate. Coordination should work and, in the abstract, there seems to be no alternative; either the authorities of the countries coordinate or some countries have outcomes they would prefer not to have, perhaps balance of payments crises and default on debt. Yet coordination has never occurred. If the authorities of the various countries can be assumed to be aware of the need, one reason for a failure to coordinate may be that some countries' authorities are confident of being able to obtain the value of output, hence the trade balances, they want, regardless of spending and production plans elsewhere. These may be the most competitive economies, which can, if their authorities so choose, have continual trade surpluses. Alternatively, a country may be able to incur debt to non-residents continually without any foreseeable need to redeem it, in which case its authorities can have continual trade and current account deficits if they wish. Another reason, outside the scope of the present discussion, is insufficient political cooperation.

To take the second possibility first, a country that has no discernible limit to its trade and current account deficits has no need to coordinate expenditure with other countries. Such a country's currency is necessarily acceptable for settling international transactions and as international reserves without limit to other countries, not necessarily all but some with surpluses enough to accumulate the supply of that currency that the others do not want to keep. Given the trade balances of the most competitive countries, this deficit country allows other countries to have more favourable trade balances than they would have otherwise and to be at or closer to the limits of their production possibilities.

This does not rule out the first reason for the lack of coordination. But the economies with surpluses would have an inducement to coordinate if there were some measure that the authorities of countries not at the limits of their production possibilities could take that would make their countries more competitive at a stroke, that is quickly increase their production, given the spending and production plans in the others. Any country that took that measure and kept expenditure down would improve its trade balance and

cause the value of production in other countries to be lower than their firms and authorities had expected; a less competitive country becomes more competitive and causes the value of production in the countries that were more competitive before to fall. If there is no country that can have unlimited deficits, at some point even the countries that were the most competitive can have falls in the value of their production. Unless their authorities judge this to be desirable, it may induce them to coordinate with the others. That is to say, if there were such a measure, its use or possibility of use could deter a refusal to coordinate.

If such a measure existed, it would be lowering the wage, which is set in terms of the country's currency, in terms of foreign exchange, either by reducing nominal wages or by devaluation, though the former can be ignored as impractical. When firms do not produce to capacity it is because the prices of the additional goods they would produce would not be greater than the additional costs per unit of current material inputs and labour. Devaluation can, then, yield a profit and allow the firms to produce more.

So, an explanation is needed as to why, despite the possibility of devaluation, there is no coordination. The question raised is, are there any *a priori* grounds for believing that devaluation makes a country that is not at the limits of its production possibilities and does not have redundant production capacity more competitive and, therefore, certain of increasing its output and improving its trade balance when its expenditure is restrained? If there are, the prospect of devaluation by competitors and the consequent displacement of domestic production should lead countries to confer.

To be realistic it has to be assumed that every tradable good is produced in several countries. Whatever the exceptions in reality, they are not important enough to warrant separate consideration here. International trade is competitive and the common argument, that devaluation allows exporters and domestic producers to sell more by lowering their prices relative to those of their foreign competitors, cannot be used. When producers in one country lower their prices of a tradable, their competitors in other countries do the same if the alternative is to lose sales and profits, or they reduce production that can no longer cover costs, perhaps stop producing altogether. Capital investment is a sunk cost and the foreign competitors continue producing as long as their profit margins gross of depreciation are not negative. Hence, for a country to increase the quantity of its exports and reduce the quantity of its imports, not only must prices fall enough to cause the gross profit margins on some production by foreign competitors to become negative, but the country must have some unused production capacity it can use immediately, for it can be supposed that the lower prices do not increase world demand enough to take up the additional production. When devaluation works, more capacity

utilisation in the devaluing country means less elsewhere; capacity redundancy is transferred from one country to another.

Therefore, devaluation may make a country more competitive, at least for a while, but there is no assurance of it. And, when it does, there may be countries to which it provides no inducement to coordinate. When a country increases its production by devaluing it may only cause countries with similar economies to reduce their production. Other countries that are more competitive or have sufficiently different economies may be unaffected and have no reason to coordinate. Hence, when one country becomes more competitive because of devaluation the others most affected may devalue in due course as well, whilst the most competitive ones have the trade balances they want. In other words, devaluation by the less competitive countries is not enough of a threat to the more competitive ones for them to want to agree on adjusting expenditure.

Other arguments, apart from increasing competitiveness, have been given as to why devaluation should result in an improvement of the trade balance and, since other countries' trade balances are necessarily worsened, the threat of devaluation can still induce countries to coordinate their expenditure. Such arguments must work through the change in the prices of tradable goods relative to some quantity internal to the economy, of which the three that come in question are the composition of costs, some form of wealth, such as the stock of a financial asset, and the prices of untradable goods.

Cost can be decomposed according to types of incomes, wage, profit and rent, or according to types of goods, in which case it would be the prices of tradables and untradables. There is no reason why changes of the composition of costs in terms of profit, wages and rent should have a predictable effect on the trade balance. Alexander points out that the higher profit rates resulting from devaluation can stimulate investment, increase expenditure and worsen the trade balance,¹¹⁰ though it is assumed here that the authorities can restrain expenditure. Less conventionally, it can be asserted that higher real wages or higher profit rates lead to more productive workers or more investment in productive capacity and R&D, but any such effects are neither certain nor immediate. Rent is the cost of space and is discussed as an untradable under that heading.

Arguments relying on forms of wealth cannot refer to production costs in the short run, though they can, in principle, repeat the assertions just mentioned about productivity of workers and investment. Those models or theories that have been put forward rely on forms of wealth working through the effects on expenditure and are, therefore, only ways of limiting the

¹¹⁰ Alexander, "Effects of a Devaluation on a Trade Balance," 273.

determination of spending that need not be discussed further. Of these the best known is the monetary theory of the balance of payments, which has long been a standard part of the courses the IMF gives for officials from low wage countries and is also taught in textbooks. In all such theories or models devaluation is assumed, not demonstrated, to result in an improved trade balance and they need not, therefore, be considered further.

Untradable goods

Arguments relying on the ratio of the prices of tradables to the prices of untradables have also been standard in the same IMF courses and textbooks. They seem to provide mechanisms for improving the trade balance by devaluation different to that described above and which would, therefore, be another way of inducing countries to coordinate. But they do not actually do so, for in all of them the stated effects of exchange rates on the price ratio and the trade balance are assumed, not deduced. Moreover, none are complete, for they make no reference to other countries or to saving and investment.

Proponents of these arguments assert that, since devaluation raises the prices of tradables relative to those of untradables, expenditure is 'switched' from tradables to untradables and that the balance of trade improves as a result. Dornbusch, for example, defines the real exchange rate as the ratio of the prices of untradable goods (which he calls home goods) to the prices of tradable goods and, describing the consequence of an overly high price of the domestic currency relative to other currencies, says, "An increase in this rate will make home goods more expensive and will encourage consumers to substitute the less expensive traded goods for home goods. Relative prices do affect household spending patterns. Consumers reallocate their budget and buy more of those goods that have become relatively cheaper."¹¹¹ Fischer says, "A successful real devaluation raises the prices of traded goods relative to those of nontraded goods." He adds, "Not only does devaluation switch demand, it also increases the demand for domestic or home goods."¹¹²

Prima facie, the notion that untradable goods are crucial to the trade balance is odd. Does it mean that imbalances cannot occur if there are no untradable goods? Or that the greater the share of untradables in the economy, the greater the effect of exchange rate changes? And does the trade balance of a country with a surplus equivalent to, say, 4 per cent of GDP and an untradables sector accounting for 3 per cent of GDP have to remain in surplus after the untradables have been eliminated? Besides, how does

¹¹¹ Dornbusch and Helmers, *The Open Economy: Tools for Policymakers in Developing Countries*, 82.

¹¹² Fischer, "Devaluation and Inflation," 119.

“expenditure switching” affect total expenditure and production? A closer look at the arguments confirms the doubts.

Statements such as those just quoted seem to imply that the improvement in the trade balance can be deduced directly from the switching of expenditure, but no such deduction is ever provided. Instead, when they set the argument out in more detail, Dornbusch, Fischer, Helmers and the other proponents of the argument assume, without explanation, that devaluation increases the quantity of exports and reduces the quantity of imports, or simply that it improves the trade balance. What is supposed to be the conclusion is an assumption.

What can be deduced as the consequence of devaluation is that expenditure switching need not occur and that the assertion that it does occur is superfluous. According to ordinary consumer theory, the amount consumed of a good normally falls if the price rises and expenditure does not change, but, though they rely on the theory, neither Dornbusch nor Fischer specifies the amount of expenditure. Fischer asserts, without explaining why, that devaluation results in “excess demand” and that it is necessary to “squeeze domestic demand to free resources for production of traded goods”, otherwise wages will rise.¹¹³ His concern seems to be that the switching of expenditure to untradables can cause their prices to rise and reduce the switching. Had he specified that expenditure in terms of the domestic currency was the same after devaluation as it was before, it would have followed that the trade balance improved because the higher prices of untradables raised the value of output relative to expenditure. The balance of trade improves even if spending on untradable goods is unchanged, because the prices of tradables are higher, the value of output is greater and the quantity the residents can buy is smaller. Devaluation leads to higher prices in terms of the domestic currency and, if expenditure does not change, to an improvement of the trade balance. There is nothing in the argument as to why simply cutting back expenditure should not give the same result.

Helmers does say how expenditure changes. He starts with a country that has an exchange rate of Rp.5 to the dollar and is running a trade deficit of \$100 million a year, whereas its trade, it is assumed, would balance if the exchange rate were Rp.7 to the dollar. When financing for the deficit stops total demand falls by \$100 million. Assuming that the country devalues to Rp.7 to the dollar and that the fall in demand for exportable goods is \$25 million and for imports \$35 million, demand for untradables falls by \$40 million. Untradables are, therefore, in excess supply unless their prices rise and their output falls, which happen because the switch of expenditure raises their prices in terms of the local currency and the relative rise in the prices of

¹¹³ Fischer, 119–20.

tradables reduces their supply. It follows that "no explicit policy is required to cut back expenditures"¹¹⁴.

But Helmers ignores the equilibrium that comes from transferring workers from producing untradables to producing tradables without devaluing. If, at the old exchange rate a fall in expenditure results in a fall in the demand for untradables of, say, \$70 million and of tradables of \$30 million, both trade and demand for goods can be kept in balance by reducing the production of the former by \$70 million and increasing the production of the latter by the same amount. Expenditure is lower, but is not switched. This may be the only solution, for Helmers does not complete his argument to show that his own proposal results in equilibrium.

The assertion that devaluation necessarily causes the prices of untradables to fall relative to those of tradables is inconsistent with competition, for it assumes that the rate of profit and other non-wage earnings do not adjust to become equal for tradable and untradable goods. If these profit rates were uniform beforehand, they are no longer so, but are higher for tradables. If they were to become uniform again after devaluation, there would be no reason that the prices of untradable goods should fall relatively. They could even rise, as can be seen from the following example.

A unit of a tradable good is assumed to have a local currency price of 100, which is the world price at the going exchange rate. Of this 30 is labour cost, 60 is the cost of inputs other than capital, which are assumed tradable, and the rest is profit. After devaluation the local currency price is 125. Since labour costs do not change and the cost of tradable inputs rises to 75, profit goes from 10 to 20. Assuming for simplicity that all the capital goods are tradable (and leaving aside the complications of amortisation), the rate of profit rises by $[20/10]/(5/4) - 1 = 3/5$.

Electric power is assumed to be untradable and to be generated thermally only. Its unit price is taken as 100, of which the labour cost is 20 and the cost of inputs other than capital, which are again all tradable, is 60. Assuming again that all capital goods for producing electricity are tradable, for the rate of profit on capital to be the same, profit must rise from 20 to 40 and the price of electricity must rise to 135. Therefore the price of electricity rises in terms of foreign currency and relative to the prices of tradables.

Objections to this reasoning must rely on market inefficiency, such as delays, perhaps of several years, in the rise of profit margins in the production of untradables. A temporary gain, then, is gradually reversed as competition

¹¹⁴ Dornbusch and Helmers, *The Open Economy: Tools for Policymakers in Developing Countries*, 15.

reduces differences in profit rates, and the prices of untradables rise higher relative to those of tradables than they were at the start, leaving the question as to how to cope with the trade deficit that, according to the standard argument, occurs as a result. On the other hand, if the reasoning is accepted and profit rates are assumed to adjust promptly, devaluation may be unnecessary; if untradables are more labour intensive than tradables, their prices can be reduced relative to the prices of tradables by reducing the profit rate by raising the nominal wage.

Fallacies apart, the notion of expenditure switching, itself, is misleading. Expenditure on tradable and on untradable goods are not necessarily distinct; all prices to the final users of all tradable goods include an untradable component that is not directly affected by the exchange rate. It was pointed out in Chapter 2 with respect to the final prices of imported branded, differentiated goods. To take the simplest example, that of an imported consumer good, the difference between the border price and the price in a shop includes port handling charges, storage and, perhaps, finance and insurance charges before transport to the shop. These are often the costs of the wholesaler or middleman. When sold at retail, the price must cover the cost of the people employed by the shop, the rent of the premises, electricity etc., insurance and often packaging or wrapping. In addition there are indirect taxes and fees charged by the state. Competing goods produced in the country include similar costs in their final prices. In a restaurant the price of a cut of meat that was originally imported includes various similar charges before the restaurant's costs of rent, fuel, storage, sanitation and staff, including cooks, waiters and scullery. Even a single person selling from a roadside stall has to have a margin to make a living and to pay for the right to keep the stall where it is. The retail mark-up of some consumer goods can be 100 per cent and the border price of an import, therefore, less than half the final price. Exports have similar costs between the factory gate price and the border or f.o.b. price, though usually fewer and smaller.

Conversely, as the example of electricity above shows, the prices of untradables include costs of tradables. Switching from one type of good to the other does not mean that one type of expenditure has been substituted for the other. Rather, if devaluation lowers the prices of untradable goods in terms of foreign exchange, it lowers a component of the cost of tradable goods too, and the higher border prices of tradables raise the prices of untradables in terms of the domestic currency.

Exchange Rates and Prices

It may seem obvious that changes of exchange rates must affect the final prices of goods, but neither the mechanisms nor the speed follows simple rules. To discuss them it is assumed here that the law of one price (LOP) holds for homogeneous goods. Branded, differentiated goods, which for brevity will be referred to simply as branded, may be but are not necessarily priced to market. All goods' final prices have an untradable cost component and the LOP applies either to border prices, the prices at which imported goods are delivered at the border, before incurring any local costs or charges, or to the prices at which exports are delivered for transport abroad. It matters little which. For simplicity the discussion is confined to devaluation, or depreciation, and, therefore, to rising domestic prices, though it is not meant to explain inflation in general but merely to show how exchange rates should be taken into account. Inflation has here the conventional meaning of rising prices of goods and non-factor services and excludes the prices of financial assets and nominal wages.

Changes of exchange rates affect domestic prices through two mechanisms. One is changes of border prices, to be considered here first, and the other is the effects on income, considered after. It follows from the LOP that the border prices of homogeneous tradables in terms of the domestic currency rise at once with devaluation, except when the price of a good is fixed in terms of that currency. Higher prices of homogeneous, tradable current inputs, which raw materials to some extent are, raise production costs other than profit. So do inputs in the form of intermediate goods, which are mostly branded, to the extent that their domestic prices rise.

Effects of border price changes

How much border prices of branded goods rise in terms of the domestic currency depends on the goods and the circumstances of the country, but they do not fall. In big industrial economies imported branded goods that compete against a large domestic output are priced to market and devaluation causes their prices to rise little if at all in the short run. Economies like Germany, Japan and the US produce more different kinds of branded goods than do smaller ones like Finland, Norway and Switzerland and, so, larger proportions of their imports are branded goods priced to market and their prices are less immediately affected. Conversely, the prices of imported branded goods rise more quickly in the smaller high wage countries. They can be expected to rise fastest in low wage countries, since these countries rarely produce their own brands, though many make components or entire products for foreign brands. In terms of the currencies of the exporting countries, the prices low wage countries pay for imports of

branded goods are unaffected by the exchange rate and the final prices in these countries rise correspondingly fast.

Little more can be said *a priori* about the pace and extent of price increases in the short term. They vary according to the good, the size of the competing domestic industry, perhaps the proportion of its output that is exported and some of the characteristics of the economy. Detailed study of specific industries and empirical work may provide some rules about how prices behave, but simple theory does not.

Beyond the short term, part of the process is the growth of the difference between border and final prices as the effects of higher border prices on the costs of untradables, like rent, transport and energy, spread through the economy. The effects on prices of transport and energy are obvious and need only be mentioned. But rent, the price of space per unit of time, is usually ignored as an insignificant part of cost, though it constitutes a large part of the final prices of goods, especially of consumer goods and, therefore, influences economic policy by being a determinant of the various price indices used to assess the effects of exchange rates on prices and to compare the price movements of different countries.

Rent is not directly influenced by exchange rates but is likely to increase in nominal terms along with price increases, though unevenly according to place. What is commonly termed rent should be notionally divided into payment for the use of buildings and infrastructure, which are produced and are capital goods, and payment for space, rent proper. Profit and depreciation on the cost of buildings and infrastructure are part of the untradable costs already included. But rent proper depends on location, since space in any location is not produced, and also on income and its distribution. Land in cities, around deep water ports or in popular seaside resorts fetches rent according to its location, in contrast to farmland, the value of which is determined by its crops. Just as the rent of farmland is determined by its fertility and prices of crops, the differences of rent between the centre of a big city and the outskirts are the consequences of location and income and its distribution, and the cost of industrial land the consequence of location and profits, whilst seaside prices may depend on the incomes of tourists. To the extent that devaluation causes prices of crops and nominal incomes to rise, it increases rents.

In principle, another reason for prices to rise is that devaluation should lead to higher depreciation charges to allow for the higher expected replacement cost of capital equipment. Even if imported capital goods are priced to market, as domestic prices rise their prices eventually rise, too. Depreciation charges and, therefore, gross profit margins should be higher, though, in practice, they are not adjusted to the exchange rate.

Income effects

The second mechanism by which exchange rate changes affect domestic prices is through incomes. In addition to the effects of changes of border prices, changes of final prices, including the costs of untradables, can be decomposed into changes of profits and rent, for nominal wages are assumed not to change. Profit rates rise after devaluation, but do not rise uniformly in the short run. One reason is that part of the tendency to uniformity of profit rates comes through withdrawal from production where profit rates are lower and investment where they are higher, and both have costs and take time. In the former capital equipment may be scrapped earlier than originally expected, though still usable, and sunk costs written off as lost. In the latter there has to be investment in new production capacity. Further exchange rate changes can add to the lack of uniformity and can give rise to expectation of more, even unendingly, and thus affect the decisions firms and individuals make about investment, consumption and in what forms to keep financial assets and other forms of wealth.

Devaluation and inflation

When a country's currency is devalued the authorities and international financial and aid organisations try to hold back the consequent inflation both because they dislike inflation *per se* and because they believe it cancels the improvements of the trade balance that devaluation is intended to bring about, and they do this by restraining expenditure. In the bigger industrial economies expenditure restraint is likely to affect the increase of the factory gate and border prices of branded tradable goods only to a minor degree in the short term. Domestic competition is likely to prevent the prices of branded tradables from rising immediately, even though domestic products would not be less competitive against imports if their factory gate prices were to be raised to offset the devaluation. Devaluation would allow firms not producing to capacity to increase their production of branded tradables if they could raise their factory gate prices, though it is still likelier that the competition of domestic producers prevents it. Hence, there is little gain from increased profit margins in the short term. Less or more expenditure means more or less exports and fewer or more imports of these goods, but the main immediate influence on price is the domestic competition.

In low wage economies and small high wage ones devaluation causes the prices of tradable goods to rise immediately and the costs of untradables to rise faster than in the big high wage countries. But wages are lowered and that is sometimes believed to make the economy more competitive. Usually this is taken to imply that nominal wages must be kept from returning to their former level in terms of foreign currencies. That lower wages make for competitiveness is an assumption as old as Hume's theory of the balance of

trade, but the assumption is taken further in more recent theories to make the trade balance explicitly a function of the wage rate. An early formulation is that of Swan¹¹⁵ and something similar is usually taken for granted when discussing how to improve the trade balance of a country. From assuming that the trade balance is worse when wages are higher, it is deduced that the trade balance is improved by lowering wages.

In practice the monetary authorities and their orthodox advisers, especially from the IMF, do not consider lowering wages relative to foreign prices to be enough for low wage countries with balance of payments problems. They almost always require such countries to devalue but they also require what they consider to be “real” devaluation, by which they mean that domestic prices should also be lowered in terms of foreign currencies. Most often they gauge prices by the consumer price index (CPI), the wholesale price index (WPI) or the GDP deflator, the rent component probably being biggest in the first, which is the most used, and smallest in the last, the least used. There is real devaluation if any rise of the chosen price index is less than the rise of foreign prices. From the chosen price index follows the real exchange rate (RER), the change in the price index in terms of the other currencies calculated by dividing the proportionate change in the price index by the proportionate increase of foreign prices, also given by some index, in terms of the domestic currency. Thus, if the price index has increased by a smaller proportion than foreign prices, there has been real devaluation.

Real devaluation is an ill defined criterion for determining the exchange rate, for it is a comparison of changes of index numbers and should, therefore, be comparison with a state of affairs that is for some reason a suitable base for it. Unless there is sound reason for choosing that base, though usually none is given, the comparison cannot be trusted. Thus, if the authorities judge that the trade balance has to be better than it was in a particular year, they can want real depreciation in comparison to that year. But, if that year was exceptional, perhaps because inflation had been especially fast or because something outside their control had made for a worse trade balance, the authorities would do better to choose a more normal year as base.

It is, moreover, hard to say how the domestic price index will have changed when the inflation caused by devaluation is over, even assuming constant nominal wages. Border prices of tradables remain constant in terms of foreign currencies, but the lower wage can reduce the costs of untradables and final prices. There is also, as shown earlier, the theoretical possibility that higher profit margins raise the prices of some untradables by even more than the devaluation. Rent changes in complicated ways; the redistribution

¹¹⁵ Swan, “Economic Control in a Dependent Economy.”

of income affects the prices of different types of urban land differently and higher profit margins may alter the relative prices of industrial and agricultural land. Since untradables, rent in particular, constitute part of the final prices of goods, the change of the price index is the outcome of a complex process and cannot be reliably predicted without estimating these various elements, though no such exercises seem to have been undertaken.

One consequence is that making a real devaluation permanent can result in a self-perpetuating process of devaluation and inflation. If the authorities want a real devaluation exceeding that which would result when the inflation caused by devaluation ends, or if they do not appreciate that price rises caused by devaluation take time to be completed, they will after a while believe that the inflation is reducing the competitiveness of the economy. They devalue before the price increases consequent on the previous devaluation have been completed, each devaluation to offset the inflation from the previous ones adding more inflation. If, instead of only comparing the movements of price indices, that is to say the RER, the authorities were to make direct, detailed comparisons of domestic prices to final buyers, they would find that, allowing for taxes, these prices are on the whole lower in countries that have devalued often in the recent past than in countries that have not. It is already evident from the rough comparisons afforded by RERs and would suffice if the authorities and their advisers did not want more.

A currency can also depreciate, not because its authorities want it to, but because its supply causes its prices in terms of other currencies to fall. If the country's currency is not accepted by others for settling international transactions the authorities, usually the central bank, can either provide foreign exchange to importers at a fixed price in terms of the local currency or provide it through some market mechanism. The quantity of foreign exchange available at any time is determined by the country's exports, net factor payments including remittances from individuals living or working abroad, net foreign borrowing, net inflows of investment, etc.. Such power as a low wage country has to influence these in the short or medium term is limited to hastening or delaying some of the payments by creating expectations about its exchange and interest rates. Hence, the supply of foreign exchange is given.

But the demand is determined by the expenditure of the country; it is the excess of total expenditure over spending on domestic production that is not exported. If the authorities set official prices for the various currencies constituting the foreign exchange supply and the demand exceeds what is available at those prices, they can provide it at the official price to those who want to buy it according to some procedure and leave some of the demand unmet. Or they can eschew official prices and use a market mechanism,

such as auctions or a regular market, in which case the prices will bring demand to equal supply. Then, the prices of foreign exchange can be repeatedly increased by expenditure. Depreciation occurs through a market mechanism, though the determinant of the exchange rate is the expenditure decided by the authorities.

When there is unmet demand at official prices, there is likely to be a black market in which the determinants of foreign exchange prices are the excess of total expenditure over domestic unexported production and the foreign exchange bought, the risks, costs and penalties associated with such transactions and the amount of foreign exchange that is available for that market.

Inflation and stability

Since untradables cannot be conveyed over their country's borders it is to be expected that some cannot be conveyed to different parts of the country. In particular, space has a location and is not produced. Rent is, therefore, not directly affected by the exchange rate and is only limited at any location at any moment by the availability of inferior alternatives, as described above. But it is influenced by expenditure. An increase of expenditure with exchange rates unchanged increases demand for types of space like those just described, as well as raising the value of production and worsening the trade balance. With that it raises property prices. Higher prices or rents for immovable property, especially urban land, raise the costs of office and retail space. In the end, goods commonly thought of as tradable cost more because they include rent. Consumer goods and, therefore, price indices, have the largest component of rent and are the most affected.

Additional demand for commercial and residential land can come from two sources other than increased expenditure. One, which will not be taken up here, is foreign demand, foreigners investing in urban property because they like the location, as an investment, to escape taxes or for other reasons. The other is the use of immovable property, land and buildings, as collateral for loans. When the authorities allow expenditure to increase, some of it is in the form of loans by banks and other financial institutions to firms and households and backed by property. Monetary policy allowing more expenditure provides the means for more loans, of which part may be spent on residential or commercial property. If property prices rise and borrowers and lenders interpret that as greater wealth and a higher value of collateral, a self-perpetuating increase of credit can ensue if the authorities allow it. This need not be limited to property; the prices of shares on the stock market, which have not been included in the discussion so far because they are not goods, serve in the same way. In principle, property and share prices can be bid up even though the authorities do not let total expenditure on goods grow. Some people or firms use their financial assets or obtain loans to buy

property or shares and those who sell reduce their debt. A purely speculative bubble can follow without aggregate expenditure on goods increasing, though one that is likely to end quickly.

When rising prices of property and shares increase their wealth, the owners may increase their spending more than any increase of income from that wealth if the authorities allow it. Rational or not, such “wealth effects” occur often. Then, the increased spending is financed by the domestic financial institutions and when spending increases beyond income from production and net factor income from abroad, the difference is financed by borrowing from non-residents, including paying with the domestic currency, which is an obligation of the monetary authority, or by selling assets to non-residents or by foreign investment in the country. If the other countries do not keep their total spending below the total value of their production and all countries are at the limits of their production possibilities, inflation ensues. If there is no inflation, aside from property prices, some countries are restraining their spending or are not at the limits of their production possibilities. In either case those residents of the country who borrow to spend more than their incomes and who devote part of their spending to property or shares increase the value of their collateral and their wealth, though that spending goes to others who realise capital gains.

As households and firms accumulate debt they can reach a point at which they become unable to service their debt if interest rates rise, except with more loans or by a reduction of expenditure on goods. In addition, as the loan demand of the more creditworthy borrowers is met, financial institutions lend to the less creditworthy. Some of this may be concealed by the rise of property and shares prices from the lending, so that financial institutions, concerned only with their own clients, believe their loans to be backed by adequate collateral and do not see that a fall of prices can be widespread because of the extent to which debtors become unable to roll their debt over.

Were the authorities to stop the increase of debt or to raise interest rates at this point, the outcome would depend on how a reduction of expenditure affected production. To take the extreme case, that the trade balance does not improve in terms of the country’s currency, the reduction of expenditure by those who needed more loans to service their debt means an equal fall in the value of production and can set off a classic crisis, the self-propagating process of falling expenditure and production, failures by households and firms to service debt and falling prices of property and financial assets. This is not a balance of payments crisis; a hard currency can stay acceptable as a means of payment, i.e. remain hard.

Here the assumption, that the authorities can determine expenditure, has had to be modified. It is still valid when expenditure is to be restrained or reduced; the monetary authorities can put limits to lending by means such as increasing the reserve requirements of banks, open market operations and direct instructions, and they can increase the cost of borrowing. These are constraints and prevent lending from increasing or make it too costly. But it does not hold in the reverse, allowing more lending and lowering the cost, which is permissive; it allows firms and households to increase their borrowing and expenditure, but does not make them do so. Some households may not want to do so, for instance, because they have as much debt as they consider prudent. If a crisis occurs following a period of excessive borrowing and spending it may be impossible for these households to avoid reducing their spending. Debt that seemed prudent beforehand can have become too costly because interest rates have been raised. Wealth in the form of shares and houses may have been reduced because the crisis has reduced their prices and, where this does not induce creditors to call in loans or demand more payments, it is an inducement to save more to improve the ratio of wealth to income. Crises result in losses of employment, which both reduce income for spending and induce people to save more as a precaution. Firms producing for that country, especially ones producing untradable goods, have to find alternative markets and there is *no a priori* reason that they can, at least in the short run. If they cannot, they reduce investment and further reduce expenditure.

Once the possibility of crisis is there, the authorities may be unable to prevent one with their normal means, though they be aware of the danger and avoid restricting spending. They can avoid raising interest rates and, if the economy is big enough and its currency is hard or its debts to non-residents are denominated in its own currency, the effects on it of interest rates rising elsewhere will be mitigated, assuming that capital movements are minor. They can also let financial institutions accommodate increases of debt service caused by higher interest rates with more loans. But none of this does more than avoid setting the crisis off, and one can still start spontaneously. Eventually the amount of debt that residents are unable to service rises or expenditure on goods falls. A crisis can also start earlier than that once enough of the country's firms become aware of the possibility; a minor, accidental event, such as the failure of a well known firm, or a rise of interest rates can set a crisis off. The central bank may even be obliged by its own precepts to raise interest rates and stop the growth of spending and, thus, provoke one. Also, there can be capital flight and the currency may depreciate because the market expects it to, which sooner or later causes prices to rise. For the central bank rising prices are reason to raise or accept the rise of domestic interest rates and usually to restrict lending, and so are higher foreign interest rates. Higher foreign interest

rates also raise the debt service by residents in terms of their currency and add to the current account deficit.

Instability of this sort does not occur when the country's residents are prudent in their borrowing, or when financial institutions have self-regulating mechanisms that stop them in time, or when the authorities follow some rule that does the same thing. In the last case, if the central bank is the authority in charge of regulating aggregate expenditure, the rule may be simply one of keeping total credit within bounds; in the words of one head of a central bank, McChesney Martin of the US Federal Reserve, 'to take away the punchbowl just as the party gets going.' It does presuppose that the authority regulating aggregate expenditure, usually the central bank, is willing to be unpopular, which may not be true of even the most independent central banker. But if, instead of these restraints, the authority considers its primary task to be keeping inflation low, as is usual nowadays with central banks, and believes its main tool for that to be some rate of interest, spending may be allowed to grow as long as a stable exchange rate or appreciation of the currency keeps inflation low, regardless of property and share prices. Low inflation can even induce the central bank to let expenditure grow by buying more debt, instead of inferring the danger of a crisis from the persistent current account deficits.

At the opposite extreme, when an improvement of the trade balance offsets the fall of spending by the residents of the country, there need, in theory, be no crisis. If the country is competitive enough it can displace production in other countries and, if it is not that competitive, countries can coordinate to increase spending to keep its production from falling. In practice, a crisis is almost certain, no matter how competitive a country may be. Crises progress fast and any form of coordination would have to be arranged in advance, for there is rarely time for the authorities of several countries to negotiate an agreement. A competitive country can be expected to recover from a crisis after a while; other things being equal, the more competitive it is the faster the recovery. If the country is not especially competitive, it can try to become more competitive by devaluing, but, if that fails, its production stays below the possibilities until spending in other countries increases enough or its production possibilities shrink as unused capital stock becomes unusable, for, as argued earlier, other countries' authorities do not necessarily think they ought to increase expenditure for some less competitive country to reach its production possibilities limits.

If countries accounting for a large part of the world's production have had crises of the sort described at roughly the same time and are producing below their production possibilities limits, i.e. when recession or depression are widespread, the other countries are unlikely to be able to provide the expenditure to stimulate their recovery. In particular, if the other countries do

not have hard currencies and adequate reserves, increases of expenditure by them can result in a loss of what reserves they have with little effect. Those that have debt denominated in hard currencies must at least obtain some form of postponement on favourable terms of their debt service as part of any agreement to increase their spending. When a hard currency country that is less competitive increases its spending, much of the consequent increase of production, aside from the increased output of untradables, is abroad. It may also put its status as a hard currency country at risk, in which case it must allow for the possible loss of its reserves. If many such countries were to coordinate to increase their expenditure, they would generate more production, beginning in the most competitive countries, which would reach the limits of their production possibilities without increasing their expenditure.

So, recovery is best started with more spending over several periods by the most competitive countries and by the countries with the hard currencies that are least likely to lose their status of being hard. The former reach the limits of their production possibilities first, if they are not already there, and probably also have ample reserves. Those that increase spending despite not being especially competitive may have worse trade deficits at the start and must either have hard currencies or enough reserves.

But, even if the monetary authorities of all countries agree on coordinating to increase expenditure, there is no assurance that they can bring about an increase. As argued above, households and firms may prefer to pay off debt and add to their financial wealth before consuming or investing more. In less competitive countries they may be even more reluctant to spend if production there is being reduced as more competitive countries improve their trade balances. Some banks will have made losses from a crisis and be unwilling to lend to any but the safest borrowers, which often means calling in other loans, and they may be constrained by having to restore their capital and reserves according to the prudential regulations and their own judgement of their needs. The authorities can help them over these constraints by providing special financing or waiving the rules for a while. Then the central bank can provide liquidity by increasing the money supply by buying securities of various sorts from the public, especially financial institutions, what is usually termed printing money, and it can lower interest rates by reducing the interest rates it uses to determine the lending rates of financial institutions. Even then neither households nor firms may want to increase spending. As Rist pointed out in 1933 and has been repeated by many economists since, lowering interest rates and allowing more lending do not result in more spending on consumption or investment for there may be no demand for such credit. Keynes compared these methods of trying to increase expenditure to 'pushing on a string', yet they are the only orthodox ones a central bank can use.

Countries do not coordinate and, as has been seen, it is the most competitive that have the least interest but whose cooperation is the most needed. If the central banks of the most competitive countries believe that increasing the money supply to increase expenditure is inflationary and that inflation must be kept low, they will not provide the necessary addition to spending. Since these countries usually have trade surpluses, it amounts to believing that balanced trade is inflationary. They can go so far as to keep unemployment up to prevent real wages from rising in the belief that nominal wage increases are one of the main sources of inflation, if not the principal source, even when they are below inflation. They may regard trade surpluses as economically desirable, though they are indefinite, free loans of the country's income in the form of foreign currency reserves.

In this case the main increase of spending must come from hard currency countries of which the authorities are not averse to trade and current account deficits and do not think such deficits will put their currencies in question. More spending by these countries increases production in the most competitive countries first, unless the authorities of these prevent it by corresponding reductions of expenditure or they are already at their limits. This may, however, be moderated in some of these countries, especially those with the biggest trade surpluses, if their authorities decide to reduce the surpluses, i.e. add to spending to that extent. The countries increasing their spending and running current account deficits benefit by consuming and investing more at little or no cost to themselves, but rather at the cost of the surplus countries. But their spending does not necessarily increase, as already pointed out, and the liquidity created by the central bank then just accumulates in the banking system. Share prices may rise and the price of urban land may too, in part because financial institutions, especially pension funds and insurance, can be obliged to invest, and the authorities can hope that a wealth effect will stimulate consumption. But, for this, households would have to borrow or slow the reconstitution of their financial wealth and may be reluctant to do either on the strength of higher share and property prices if they have just been through a crisis in which these prices fell. Moreover, if the monetary authorities are successful there is the additional concern of the growth of debt, as before the previous crisis. Success would result in bigger current account deficits, which should be used by the central bank as an indication of accumulating debt within the economy.

When an economy is not at the limits of its production possibilities, in recession or depression, the authorities can try to stimulate it by increasing the state's expenditure. Their scope for doing so depends on whether or not the central bank is independent and their success on the competitiveness of the economy. They can try to create employment directly, e.g. building infrastructure, or indirectly by increasing income through wage increases,

especially for government employees, and through transfers like pensions and unemployment benefits. Some of the employment created may be abroad, but the first method creates a known minimum. Presumably the budget is not in surplus, so the state must borrow from the central bank or from the public. Loans from the public have, in theory, to be repaid and, if the central bank is independent, its loans to the state must be repaid too. If the loans' purpose is achieved and production stimulated, it may be possible to repay them early from the tax revenue generated. But the state may have accumulated debt over several years, especially if the current account has been in persistent deficit and budget deficits have been among the causes. In that case the central bank and the public may be reluctant to provide the financing and interest rates may rise. Another possible obstacle is law putting a maximum to state debt. There is no cause for apprehension if production is sure to increase adequately, but there is if the trade balance worsens, instead, and leaves more debt.

A central bank does not have to be independent; it can be obliged to follow the instructions of a ministry, usually the ministry of finance, and it can even be a mere department of the ministry. Most central banks in high wage countries were not independent before the 1990s. With sufficient state control over the central bank the state's debt to the central bank becomes an accounting relation internal to the state's own finances that need never be repaid or appear in the state's budget. In contrast to state debt that has to be repaid, it is of no direct interest to financial institutions, non-financial firms or households. According to those who favour the independence of central banks, this is likely to result in excessive money creation. They argue that the politicians who control any state, democratic or not, have their own motives for increasing spending and, hence, for increasing the money supply, which is not normally be good for the economy as a whole and causes inflation. According to them, only if the central bank is independent can its decisions be free from political influence and there be some assurance of sound monetary policy, especially a guarantee against hyperinflation.

History does not bear out this argument. Germany's central bank was independent when the hyperinflation of 1922-3 occurred. During the 1950s and 1960s, when few central banks were independent, high inflation in high wage countries was rare. Only in Japan at the time of its fastest growth and high levels of protection was there much inflation. In Europe inflation was seldom outside the accepted bounds. But economic growth was faster and unemployment rates lower than they were to be later. After that, the only time when inflation was considered too high was the period of the 1970s, when the US expenditure on the war in Vietnam and on its social programs turned the trade surplus into a deficit and the other high wage countries had little unemployment. As was to be expected, inflation was worldwide and

also unavoidable unless the rest of the world reduced expenditure to accommodate the US, which some countries did a bit in the hope of reducing their own inflation. The US Federal Reserve was independent but did not prevent the growth of US spending. Similarly, spending grew rapidly in the 1990s and share prices rose with it, to the point that in 1996 the head of the US central bank, Greenspan, the Chairman of the Federal Reserve Board from 1987 to January 2006, expressed concern over 'irrational exuberance'. Again the Federal Reserve did not curtail the spending and there was a fall of share prices, the bursting of the "dotcom bubble" in 2000, as many commentators had predicted. Spending in the US was allowed to increase rapidly again until the trade deficit almost reached \$800 billion in 2007, as compared to deficits that had never reached \$200 billion before 1998.

That big a trade deficit should have been taken as a warning of excessive credit growth within the US. But, by 2000 the illusion that there could be no financial crisis, as opposed to a passing fall of share prices, had become common. Not only had there not been one in the period since the Second World War, but there were also new economic theories that supposedly proved there would be none. They were not theories intended to explain why crises had not occurred in the previous years, but, by relying on assumptions about behaviour and institutions that did not apply to that period, in particular with regard to financial markets, they purported to show how markets could be more efficient and prescient. An empirical test had to be in the future. Nevertheless, the financial authorities could see that in both quality and quantity lending in the late 1990s and after 2000 was growing more and more dangerous. Under similar circumstances in the years just before the 1929 crash of the New York stock market the Federal Reserve officials was already aware of the danger, but by then they dared not take strong enough action because they knew that a crash would follow if they did and that they would be blamed. Greenspan asserted in 1966 that the speculation that led to the crash of 1929 and depression was the result of the growth of credit the Federal Reserve had allowed, in particular to help the UK by keeping interest rates down. Until the crisis began in 2007 central bankers had the approbation that comes with a long financial boom and, as after 1929, they avoided the odium of the stock market crash and fall of production. Not having been elected and not having been called to account for their results like ordinary politicians, the central bankers have kept their independence and control over monetary policy.

2. THEORIES OF THE TRADE BALANCE AND EXCHANGE RATES

This discussion has so far been an attempt to draw conclusions from first principles and some of the relations of the national and external accounts of

countries. Institutions have not been specified beyond the assumption that the authorities can determine expenditure. What follows is an attempt to describe briefly how theories about the balance of payments have evolved and institutions have varied, going from the gold standard through the disorder of the years between the two World Wars to the fixed exchange rates of the Bretton Woods system and the floating exchange rates that replaced it in the 1970s.

The Gold Standard

Ricardo's theory of the balance of trade was that of Hume, which had become orthodox, and it was also part of his belief that only gold was an "invariable" standard of value and that paper money should be fully backed by it, the Bullionist view. Yet Britain's industrial and commercial growth through the 18th century had given rise to a financial system that functioned mainly with credit and of which currency was a minor part. With this came excesses of credit followed by financial crashes. According to the Bullionists and, later, the Currency School, these booms and crashes would not occur if changes in the currency in circulation were equal only to the changes of gold reserves.

Consequently, Britain's gold standard and financial institutions were the outcome of the interaction of economic doctrine and the needs of the economy. On the one hand, there was the aversion to giving any body powers to regulate the economy on the grounds that human judgement was fallible and that, when allowed to, people would misuse their powers for personal or factional advantage. If the currency were to be determined solely by the stock of gold, there seemed to be no reason for a Bank of England, and Ricardo, himself, described the Bank as unnecessary. On the other hand, The Bank of England was much the biggest bank, it was the bank of the government, which deposited its revenue there and from where it made payments, and it was the source of the reserves of the London banks, which, in turn, held the reserves of the country banks. Peel's Bank Act of 1844 organised the Bank of England according to the principles of the Currency School, following, in particular, Ricardo's proposal of separating the issue of currency from banking activities. From then on the Issue Department could only issue currency, its notes (then £ 21 million, while country bank notes were £ 8.6 million), to the value of its gold reserve plus a fiduciary issue (then £14 million) and had to convert them on demand into gold sovereigns without charge. This was the form that the Bank of England kept through the period when the gold standard was nigh universal until the First World War.

Britain's industrial, commercial and financial pre-eminence through most of the 19th century gave an advantage to adopting the gold standard, an advantage that increased with the number of countries that had already done

it, so that by the end of the 19th century almost every country in the world was on gold. It was period of globalisation, international capital movements and direct foreign investment, but the concern here is with theories of the balance of payments and the discussion is confined to that.

For Ricardo Hume's balance of trade mechanism was enough to show that countries are brought to their comparative advantages when they are not at them. For it follows from his scheme of two countries and two goods that, if they are not at them, one will be exporting both goods to the other because its prices for both goods are lower and that the consequent changes in both countries' stocks of gold change prices so as to make both countries produce according to their comparative advantages. Prices rise in the surplus country and fall in the other until one good can be produced in the latter at the same price as in the former. This means that the ratio of the labour used to make a unit of that good to the labour used to make a unit of the other good is lower in the deficit country than in the surplus country, for prices are proportional to the labour needed to make a unit of a good. Eventually the gold movements bring the prices of the two goods to a level at which trade balances. Ricardo left open the question as to where within the range allowed by the comparative advantages prices will be when trade balances. Mill answered it in his 'Principles' on the assumption that demand for each good is higher for a lower price. Then, as long as one country has a trade deficit the prices of its exported good is lowered as a result of the outflow of gold and the price of its imports are raised until exports and imports balance. Mill also extended Ricardo's comparative advantages and the balance of trade mechanisms to several countries and that remained the orthodox theory until after the First World War.

As a theory of the balance of payments Ricardo's theory had two deficiencies. It could not allow for the possibility that income or expenditure might affect exports or imports and it omitted movements of capital between countries. Output, income and expenditure were not differentiated and the authorities influenced nothing. Capital movements interfered directly with the balancing of trade and indirectly through the yield on foreign investment and could, therefore, prevent countries from reaching their comparative advantages.

Economists did occasionally argue that imports might be affected directly by changes of income or by transfers of money or gold, and not only indirectly through changes in price levels, but such arguments were no more than asides and did not affect the orthodoxy as developed by Mill. Among them were Mill, himself, Longfield, Cairnes, Bastable and Wicksell. They differed in their reasoning for the same reason that they did not develop it further, namely the difficulty of formulating an argument in terms of income and expenditure before the advent of national income accounting, which

only began to be developed in the late 1930s. Since the authorities were assumed not to influence anything, expenditure had to be taken as determined by the quantity of gold and, since countries were assumed, for the purpose of the discussion of international economics, to be at the limits of their production possibilities, income and production varied, if at all, only because of changes in the pattern of international trade. Hence, expenditure and income of the country as a whole, which are distinguished in social accounting as items on the opposite sides of the national accounts, were, in the minds of these economists, not distinct and had no clear relation to trade.

Ricardo and his followers avoided having to cope with international capital movements by asserting that investment abroad was deterred by the unfamiliarity and risk, as was still common in textbooks late in the 20th century. They must have felt that foreign ventures, although they had been common and conspicuous at least since the discovery of the Americas, had been too small in total to have had a noticeable effect on the balance of payments. Nevertheless, the effects of loans and transfers abroad were discussed by Ricardo, who would certainly have opposed any hindrance to movements of capital, short or long term.

Capital movements grew steadily and in the long run their yield became so large a part of Britain's balance of payments that trade deficits were normal. Ricardo could not have foreseen that Britain's exports of capital would average 5 per cent of national product in the last decades of the 19th century and 8 per cent in the last years before World War I. Britain's merchandise trade, valued f.o.b., was always in deficit from 1822 on. (Ricardo died in 1823 and Mill's *Principles* was published in 1848.) When valued c.i.f., the balance was favourable for several more years, for Britain still had surpluses on shipping, though by the late 19th century surpluses on goods and non-factor services had become rare. What mattered for the convertibility of the currency was that the yield on foreign investment normally kept the current account in surplus, i.e. Britain's exports and income from foreign investment exceeded its imports and payments on foreign capital. France and Germany, too, exported capital, though in smaller proportion, usually 2–3 per cent of national product, and they ran fewer trade deficits. The US was a net importer of capital until the end of the 19th century, but its trade was mostly in surplus.

Capital movements, that is to say financial transactions other than to make payments for goods and yields on investments, became the main determinants of gold stocks. In London, the main financial centre, the banks held the deposits of residents of other countries, often large deposits of governments, and the same was true, on a smaller scale, of the other main financial centres of Europe, notably Paris and Berlin. Confidence in the

system grew with the experience of its stability and was manifested in the use of the major currencies as reserves for international payments to avoid big fluctuations of gold reserves. When the Bank of England had to act to protect convertibility, it was not in respect to trade deficits, but to capital movements

Events before the First World War, therefore, gave no reason for questioning the standard theory of the balance of trade; the premises did not hold, so the theory could not be put to the test. It was put forward as a certainty in the report prepared for the British government in 1918 on the consequences of the war for currency and foreign exchange by a committee led by Cunliffe, a former governor of the Bank of England. The Report recommended the immediate return to the gold standard and gave a description of its workings that was for some time widely accepted, namely that gold came to or left London as the discount rate was raised or lowered, and added that a persistent trade deficit would have to be remedied sooner or later by using high interest rates to lower prices and, thereby, increase exports and reduce imports.

It is less certain now that the workings of the gold standard and why it lasted so long over so much of the world are well understood. The mechanisms and the history seem to be known in sufficient detail, but, as Bloomfield showed in 1959, the behaviour of the monetary authorities did not conform to the simple scheme of the Cunliffe Report. One inconsistency apparent at the time was that the Report did not explicitly take into account that interest rates in the rest of Europe and in the US might move together most of the time and it gave no instance of when Britain had needed to reduce its trade deficits or had tried to do so in the orthodox way. An effect of the stability brought about by the general use of the gold standard was that the economies of Europe and North American moved to a great extent together, their ups and downs usually coincided. Rist, adviser to French government, referred in 1933 to the almost perfect identity there had been of the price movements across Europe and the US and quoted the analogy made in 1925 by Churchill as Chancellor of the Exchequer, in his motion in Parliament to restore the gold standard, of the tide lifting all boats together.¹¹⁶

Understanding the gold standard is also complicated by its evolution over the thirty or more years before the war. One of the reasons for change was that the supply of gold grew more slowly than international trade or the money supplies of most countries. It was also unevenly spread among countries. In Britain the Bank of England kept its gold reserves to a minimum because they reduced profit, in contrast to France, Russia and the US. In 1913 the Bank of England's gold amounted to roughly \$ 170 million, the Banque de France's to about \$ 580 million and the gold reserves of Russia and the US to about \$ 780

¹¹⁶ Rist, *Essais sur Quelques Problemes Economiques et Monetaires*.

million each.¹¹⁷ Countries had at times to cooperate by lending some of their gold or not moving bank deposits to avoid the gold reserves of Britain or some other country falling so low as to require drastic action. The advantages of having bigger stocks of gold grew relative to the costs. A less obvious change was that the Court of the Bank of England in the years just before the war consisted almost entirely of bankers, in contrast to the Court described by Bagehot in his book “Lombard Street” in 1873, which was composed of men who had interests in industry and trade. Cunliffe, for example, was a banker, though the origin of the family’s wealth was railways, and the recommendations of his committee might have been different if industry had been better represented. The committee’s two academic members were Pigou and Keynes and the latter dissented from its report.

It can be argued both that the gold standard would have failed soon if the First World War had not ended it first¹¹⁸ and that it provided the economies on it, at least the main ones, with what was in practice a stable common currency, or, as de Grauwe put it, each ‘... became effectively an international currency’¹¹⁹. Failure was likely because of the relative scarcity of gold; unless countries acted in concert to raise its price, which was inconceivable then, the scarcity would have led to countries hoarding gold more and to rivalry in place of the cooperation. The gold standard was also close to providing a single currency, for restrictions on monetary transactions between countries were minimal and exchange rates were given by the prices of gold in the various currencies. It was stable because each country’s monetary authority was independent of the others. In theory, when gold was moved out of some countries it moved into others and the monetary contraction in the former was offset by expansion in the latter. If, instead, as seemed to be the norm, the economies of Europe and the US moved together, the gold supplies of the individual economies acted as separate constraints. What was needed was that the gold supply increase with the growth of the world economy or that the economies forming a *de facto* currency union have a substitute for gold that they could increase by agreement of their monetary authorities.

After the First World War

Only the US kept the gold standard in the aftermath of the war; all the countries of Europe had to abandon it. For most countries of Europe the period from then to the next world war the period can be divided into three. In the first their political leaders and financiers had no doubts about the desirability of returning to the system they trusted and thought they knew and

¹¹⁷ Federal Reserve Bulletin. 1920.

¹¹⁸ De Cecco, *Money and Empire: The International Gold Standard, 1890-1914*.

¹¹⁹ De Grauwe, *International Money: Postwar Trends and Theories*, 11.

of returning at the old parities if at all possible. It was a period of preparation for the return and turned out to be both longer and more troubled than they foresaw. Then came for almost every country the period from return to the gold standard to the start of the Great Depression in 1930. Countries joined and left at different times, but almost all were on the gold standard in the main period, from 1928 to 1932. One by one they left it in the third period, that of the depression, as did the US briefly. Since the intention here is to discuss theories of the balance of payments, what follows is not meant to be a history, but deals with events only to the extent they are relevant.

During the first period there were, in addition to the desire to bring back the gold standard, real economic needs, and many of the troubles were caused by failure to reconcile the two. Employment had to be found for the soldiers being demobilised, whilst the neglect of maintenance and replacement of industrial capacity and infrastructure necessitated by war production had to be remedied and what had been destroyed in Belgium and France had to be replaced. After the Bolshevik Revolution of 1917 the possibility of high unemployment among men who had been through the dangers and hardship of war caused alarm. But soon Russia had a civil war, in which several western countries took part, and the immediate threat of communism spreading was ended.

Typical of the confidence in the gold standard was the recommendation of the Cunliffe Report, that Britain return to it quickly. This confidence seems to have provided a guide in a situation that had no precedent and by doing so to have been a cause of stability. It did so through the orthodox theory of the balance of trade, which remained unquestioned, namely that the trade balance of a country improved or deteriorated as the price level of the country fell or rose relative to the price levels of other countries. If countries were to have balanced trade, therefore, the price levels in all should be the same, that is to say, there should be purchasing power parity (PPP). It was an argument formulated by Cassel to give a way to choose parities and was used repeatedly in the next two decades in discussions about what the exchange rates should be. The international uniformity of price movements that Rist referred to had been interrupted by the war; inflation in all the European countries had been higher than in the US during the war. But, since PPP became an accepted principle, when the wartime controls were removed the European currencies did not move unpredictably, but depreciated against the dollar roughly according to PPP.

Another possible cause of stability was the general desire of the politicians and financiers to return to the old gold prices. PPP had, therefore, to be restored at the old exchange rates, which meant that the European countries had to reduce their prices by the amounts that their wartime

inflation had exceeded that of the US. At the start speculators were so confident of the movement to be expected of exchange rates that they held currencies that would appreciate if they reverted to the old parities. As with the pre-war system, speculation was stabilising.

A consequence of these stabilising effects was that shortages of means for making international payments were not as acute as in the aftermath of the next world war. Gold was always accepted and the monetary authorities now held most of the gold that had been in private hands, so that they had the means for making international payments. Much of the gold had gone to the US, which had over \$ 2,200 million worth at the end of 1918, but the reserves of the Banque de France and the German Reichsbank were roughly the same as in 1913 and the Bank of England had \$ 385 million.¹²⁰ Russia seems to have been the main loser of gold, though civil war made the official numbers meaningless. Within Europe, at least, the same currencies that had been commonly accepted before were accepted again, sterling especially, despite depreciation against the dollar. When the politicians and central bankers asserted that only by going back to the old gold prices could confidence in the gold standard be restored, they prevented a problem from arising needlessly.

They still had two problems; one of adjusting their economies to the changes of the pattern of international trade and payments that had occurred and the other of having enough gold reserves to satisfy private demand in the event of a return to the gold standard as before. Britain and France had used much of their foreign investments to finance the war and Germany had lost all its foreign investments as well as its colonies. A large part of these investments had been in the Americas and had been acquired or replaced by US capital, so that there was a loss to the current accounts of the European countries and a gain to that of the US. In addition, there was the debt service on the war loans the US had made to its European allies. Within Europe there was the new element of German reparations payment to Belgium, Britain and France, to be paid in gold and in kind, which could prevent the recovery of the German economy. Production and employment in Europe depended on repairing and rebuilding industrial capacity and infrastructure, and the faster the repair the greater the trade deficits, which would have had to be financed with gold or foreign loans, mainly from the US.

But the political leaders and monetary authorities were more intent on protecting their gold stocks to return to gold than on production or employment and the US was no longer willing to lend large amounts on acceptable terms. There was much concern in Europe that more gold was needed and that there was less than before the war. Even if there was confidence immediately after the war that the old system would be restored,

¹²⁰ Federal Reserve Bulletin. 1920.

the reserves to maintain confidence would have had to be larger. It is unlikely, for instance, that the Bank of England could have maintained convertibility with the small reserve, amounting to 5 per cent of its liabilities, that it had kept before the war. On the contrary, several prominent economists and financiers had already criticised the Bank of England before the war for keeping its reserve too low. Much of the gold of the various monetary authorities of Europe had come from non-official institutions and the public, who might want to hold some again if allowed to do so, though this demand was suppressed by agreement at the Genoa Conference of 1922 by limiting sales of gold by the monetary authorities to bars and not issuing gold coin. A further need for gold arose because there were more independent countries in place of the single Austro-Hungarian Empire and each would want some gold reserves.

Investment in industry and infrastructure was kept down and unemployment stayed high. Britain had current account surpluses until its return to gold in 1925, though its trade balance was never positive then or after. Nevertheless, at first, before Germany returned to the gold standard in 1924, Britain, France and Germany lost gold almost uninterruptedly, whilst the US imported gold. In the three countries, the loss of income from foreign investments had the effect that the trade balance was no longer remote from transactions on reserve account and that the countries concerned could no longer run trade deficits as great as those before the war without losing gold or foreign exchange.

Aside from the return to the gold standard, there was the question of how much Germany should pay in reparations and in what form. Germany was the only big country left of the defeated powers since Austria and Hungary were now small countries and the empire they had ruled had been broken into several small states and parts of what had been the two countries transferred to Italy and Poland. It was still the biggest country of Europe, apart from Russia, despite ceding Alsace and Lorraine to France and losing some eastern territories. It also lost its colonies to Belgium, Britain and France. Reparations were to be made as payments in kind and in gold, beginning with the transfer of much of the country's transportable capital equipment, especially railway rolling stock. The full amount and the schedule were only decided in London in May 1921, but payments had to be made in the interim.

The total amount decided in London was the equivalent of over 300 per cent of national income and began with payments of 10 per cent of national income. Several people, such as Taussig, professor of economics at Harvard, and Lamont, a member of the US delegation at Versailles, criticised the

demands.¹²¹ Keynes, who had been a senior official of the British Treasury at the negotiations, argued in his book *The Economic Consequences of the Peace*, written in 1919, that Germany could only be reasonably be expected to pay about a third of the amount that was later demanded in London. Germany's prospects for making the payments were bad from the start. Not only were the amounts demanded economically destructive, but they aggravated the political division within the country by discrediting the German government, who had hoped to obtain better terms by being cooperative. A rescheduling in January 1922 gave an initial reduction of payments and led to an agreement on the budget, but they were not enough to leave the state the revenue it needed.

Under the circumstances it was impossible for Germany's currency to be accepted for settling international transaction, whereas the country's expenditure had to include reparations payments and normal payments for goods and services. Hence the price of gold in terms of German marks was determined by the country's foreign exchange receipts and expenditure, and the demand for foreign exchange exceeded the supply. It would not have had there been enough foreign loans or capital inflows or if Germany had used its reserves. But the risks were too high for foreign loans to be affordable and once capital flight had started in 1921 the net flow of capital was out rather than in. Unlike Austria and Hungary, Germany was prevented by the Versailles Treaty from controlling the capital account. Had the state's deficits been offset by an excess of non-government saving over investment, there would, at least, have been the trade surplus available for reparation payments. But this would have required a higher rate of saving than was feasible at the time, even though investment was low. The currency had been depreciating with respect to the dollar since 1918 and the rate accelerated in the middle of 1921. When it accelerated more in the summer of 1922 Germany's hyperinflation had begun. It grew worse when France and Belgium occupied the Ruhr in January 1923 to extract reparations by force; much of the production in the region stopped and the German government added to its expenditure by subsidising the industries and workers.

This was a self-perpetuating process. Even before the occupation of the Ruhr inflation was too fast for revenue to keep up. Because of the normal lags in tax collection and because rates that had to be adjusted could not be adjusted quickly enough, the proportion of government expenditure financed by printing money was greater the faster the inflation. These were not the conditions to expect offsetting non-government saving. Savings in financial form lost their value and both households and firms held as little of the domestic currency as possible. Such financial savings as they still had were in foreign currencies or gold.

¹²¹ Eichengreen., *Golden Fetters.*, 131–33.

Inflation stopped abruptly in late November 1923. Germany had enough gold reserves that, at the going price, they were the equivalent of 95 per cent of the outstanding money balances. Prices having risen by a factor of nearly 10^{11} from mid-1921 to November 1923, a new currency unit, the "rentenmark", was made the equivalent of 10^{12} old marks. It was then possible to fix the exchange rate for a while. The currency stopped depreciating in late November and prices stopped rising at the same time. By then the Ruhr industrialists had come to an accommodation with the Allies and had begun supplying coal to France as part of the reparations payments in kind. At the same time the Allies suspended some of the reparation payments, the German government stopped its financing of the Ruhr industries, postponing also payments for the coal, and it reduced employment in the administrative system and state enterprises, especially the railways. These measures made it possible to abide by the limits put on the loans that the new institution for lending to the state, the Rentenbank, could make, while confining the central bank to issuing notes with gold backing and to discounting commercial bills.

All of this would have been undone by the resumption of reparations payments, the increased unemployment, which had a budgetary cost and caused unrest, and high short term interest rates, which, because of the uncertainty, were around 22 per cent. Only when the negotiations of the Dawes plan were concluded with a loan to the German government in May 1924 was there the financing for the balance of payments and, therefore, for confidence enough for long term loans to Germany at 11 per cent.

Several economists at the time stated that the depreciation of the currency was the prime cause of the inflation. For instance, Helfferich, a former finance minister, pointed out that prices lagged behind the exchange rate and prices of domestically made goods lagged behind the prices of imports.¹²² Moreover, the depreciation and the price rises ended in quick succession, though the government printed money for several months more. Sargent gives the opposing opinion, that the inflation was caused by the increase in the supply of money and depreciation was the consequence.¹²³ He does not discuss why the increased money supply should not, at least at the start, have resulted in a greater trade deficits instead of higher prices nor whether the exchange rate might have had an effect on prices. Instead, he asserts that the market, being rational, stopped expecting inflation as soon as it was convinced that the necessary steps were being taken, even though money continued to be printed for several months, but, as Eichengreen

¹²² Helfferich, *Das Geld*, chap. 13. Section 8.

¹²³ Sargent, "The Ends of Four Big Inflation."

pointed out,¹²⁴ does not explain why there should have been such an access of faith after what had gone before.

Austria, Hungary and Poland also had high rates of inflation and currency depreciation and, just as in Germany, their inflation stopped as soon as their currencies stopped depreciating. In each case the depreciation stopped when the countries obtained foreign financing. Sargent explains the end of inflation in all three cases in the same way, the restoration of confidence, but does not explain how confidence could be restored in Austria in September 1922, just after Germany's hyperinflation had begun.

Once Germany had returned to the gold standard and there was confidence the authorities intended to keep to it, it obtained American loans that allowed it to make the reparations payments and to invest in restoring its industrial capacity. Without these loans the gold stock would have been used up and inflation would have returned at a rate determined by the country's spending expenditure Austria and Lithuania had already adopted the gold standard and Britain did so in 1925, which induced several other countries to do so. By 1928 almost all of Europe had followed.

But this time the gold standard was not as robust as it had been before World War I. It depended on US capital, especially loans to Germany. Countries receiving the reparations could service their debts to the US and make government expenditures beyond their domestic revenues. In theory this could have gone on as long as Germany did not appear to have too much debt. For instance, if the growth of the German economy and its trade surpluses at some point began to reduce the share of GDP going to foreign debt service and reparations and if there were confidence that this growth would continue, lenders could have felt safe. It had to be assumed that nothing would go wrong.

To many this was to be overly optimistic and was, therefore, dangerous. Keynes took up the subject of reparations again after the turmoil of ten years to warn that the payments were not feasible.¹²⁵ His argument was that the increase of exports needed for the requisite trade surpluses could only be brought about by reductions in the prices of Germany's goods, for which nominal wages in terms of other currencies, what Keynes termed gold-wages, would have to be lowered. Devaluation being ruled out by the gold standard, this could only be done through creating unemployment and, even with all possible reductions of imports, the increase of exports would have had to be so great that the necessary unemployment risked causing more turmoil.

¹²⁴ Eichengreen., *Golden Fetters.*, 146–50.

¹²⁵ Keynes, "The German Transfer Problem."

Keynes was being orthodox in ascribing changes in the amounts of exports or imports to prices and was criticised for that by Ohlin,¹²⁶ who argued that a transfer of “buying power” from one country to another would result in part of the transfer being used directly to increase imports by the latter and the rest being used for expenditure within the country, which, too, would increase imports indirectly. In the country making the transfer and, thus, reducing its buying power the effects would be the opposite. Germany had borrowed twice as much as the reparations it had paid without noticeable problems for the countries transferring their buying power, so a reversal of the process was unlikely to cause great problems either. Both Keynes and Ohlin agreed that the other countries should not hinder Germany’s exports by protection.

Had Keynes, Ohlin and their contemporaries who discussed the transfer problem been able to discuss in terms of national accounts as they were developed after the Second World War, there would have been little dispute. But national accounting was only then beginning to be developed and the relationships of production, income, expenditure and the external account had not been clearly formulated. Like their 19th century predecessors, the economists of the time could not state the transfer problem as one of macroeconomic balances, which is what Ohlin was trying to do. He asserted that the transfers could be made through an “organised shifting of demand”, though the import of the assertion was destroyed by the immediately preceding paragraph, in which Ohlin limited the transfers to deliveries in kind and conceded that big transfers of money might not be feasible. Like him, Keynes and several other contemporaries showed awareness that the demand for the goods that Germany could sell had to be created when they asked that the countries receiving the reparation payments specify what the goods were they were to receive.

Only in the *General Theory* did Keynes connect production, income and expenditure, though he confined himself to the closed economy. That this apparently simple step was in reality difficult can be seen not only from the lack of explicit awareness until then, but also from the statements of contemporaries. Rist, for example, could state explicitly in 1933 that a central bank cannot increase its credit during a depression for the simple reason that there is no demand for it.¹²⁷ Here he anticipated the *General Theory*. He also stated that a depression is caused by the public being unable to continue buying at the prevailing prices, but added he did not claim to offer a theory of how this could be.¹²⁸ A different opinion was expressed by Rueff in a comment on the discussion of Keynes and Ohlin; ‘... never in the

¹²⁶ Keynes and Ohlin, “The Reparation Problem: A Discussion.”

¹²⁷ Rist, *Essais sur Quelques Problemes Economiques et Monetaires.*, 107.

¹²⁸ Rist, 145–46.

course of the various economic transformations that occur is purchasing power lost or created, but that it always remains constant...¹²⁹, a principle of conservation of purchasing power that nobody else accepted.

Nowadays it is apparent that the transfer problem could have been solved if the countries to receive the transfers had adjusted their spending so as to have trade deficits financed by the transfers. Germany would then have had to keep its spending below the value of production by the amount of the transfer. Since it was the state that had to make the transfers, the taxation of its residents would have had to exceed the state's other expenditures by the amount to be transferred or, failing that, but assuming that the trade surplus was adequate, by borrowing from its residents.

It might have been thought that the prospect of being able to increase consumption and ease internal disputes over taxation without having to worry about eventual trade deficits would have been immediately attractive. The possibility for the peoples of Belgium, Britain and France to live better at the expense of Germany was obvious. But the politicians of all countries gave too much importance to what they were told were sound economic principles that created confidence in the financial markets to be willing to seek apparently short term political advantage by taking the path of least resistance, even if that would have made them popular; they were determined, at least at the start, to return to gold at the old parities. Prices and, therefore, wages had to be lowered to reverse the inflation of the war years, so expenditure had to be kept down and unemployment high.

As far as macroeconomic balances are concerned, the course that the European countries should have taken immediately after the war to return to the gold standard was to agree on a set of exchange rates, price gold at several times the pre-war prices and, since it was unlikely that they could have trade surpluses with the US, hope that the US lending would continue until something changed, be it attitudes in the US, Europe's ability to run trade surpluses or gold discoveries. Agreeing on exchange rate should have been straightforward, for all accepted the principle of PPP. Cassel had proposed the notion specifically for that purpose and it was a success in this respect. Disagreements would have been confined to questions of how to measure it. By deciding to return to the old gold prices governments avoided any need for agreement on exchange rates, but had to lower domestic prices. Later, when countries, like Austria, France and Germany, returned to the gold standard after periods of inflation, their exchange rates were the products of circumstances and, for the sake of confidence, not to be questioned.

¹²⁹ Rueff, "Mr. Keynes' Views on the Transfer Problem," 389.

Even then it is not certain that the depression of the 1930s would have been avoided. Despite the depreciation of their currencies, the European countries as a whole could not run trade surpluses with the US because of the latter's protectionism; the Fordney-McCumber Tariff raised the average rate on dutiable imports from 30 to 35 per cent.¹³⁰ Since most imports to the US from Europe were manufactures that competed with domestic goods, the tariffs were, at least partly, taxes on the exporters that reduced the income from the exports. US residents had little interest in holding European currencies as reserves or deposits, so the Europeans settled their current account deficits in gold or by borrowing from the US. Hence, Europe was borrowing from the US and servicing loans to it. Germany was especially dependent because it had to make reparation payments, which the recipients used to service their debts to the US, and ran trade deficits in addition. Gold payments to the US were minimal; virtually all the increase in the world's monetary gold before 1925 had gone to the US. After 1925 the US began to lose gold and until 1936 the increase went to France.

Such direct and indirect dependence on capital from the US without the possibility of servicing debt by trade surpluses or big inflows of gold made it virtually impossible to avoid a crisis when US lending came to a stop, as it began to do in 1928, which was also when signs of a depression began to be seen. Then came the stock market collapse of 1929 and by 1931 the US and most of Europe were in depression. With high unemployment and failing industries and banks, countries were leaving the gold standard. Because of the size of its gold reserves, France avoided the depression and kept to the gold standard longer than all but a few, but it, too, eventually succumbed to the depression and abandoned the gold standard. As before the return to the gold standard, countries either allowed their exchange rates to be determined by the market or tried to peg their currencies to the currency of a major economy.

New Theory

The extent of disagreement during these two decades of economic troubles over the nature and causes of the problems of the time could not have been foreseen before the First World War. But the problems had not been foreseen either; restoring the gold standard by deliberately lowering wages and prices, the concentration of monetary gold in two countries, the reparation payments, inflation and hyperinflation and unemployment caused deliberately at first and then by the depression were problems the orthodoxy did not admit and for which it did not have an answer. Some questioning of orthodox economic theory followed, although, as regards international economics, it was confined to the theory of the balance of trade. Ohlin's

¹³⁰ Eichengreen., *Golden Fetters.*, 105.

exposition of the factor endowments theory, though published in 1933, was not so much part of this questioning as the outcome of what had begun with Heckscher in 1916, but his attempt to link the balance of trade to income arose from the German reparation payments, and he was followed by Harrod and Joan Robinson in trying to formulate the link. What remained unquestioned was that, other things being equal, raising and lowering the prices of one country in terms of the currencies of other countries reduced or increased the quantity of exports from the former to the latter and vice versa for imports. In principle, therefore, to the extent that prices changed with them, changes of the exchange rate or the wage rate could be used to increase or decrease exports and imports. Prices in one country might also change in terms of the currency of another because of inflation or deflation, either of which could be offset by a change in the exchange rate. At the same time, it was generally accepted that devaluation caused inflation and that was sometimes a deterrent to trying to improve the trade balance by this means.

Since exchange rates could now be changed at the will of the authorities, or be left to be determined by the market, there was no need to be able to change the money wage, especially if to lower it. Eventually economists began to assume that nominal wages could not be reduced. One reason was Keynes's argument,¹³¹ that lowering money wages did not increase employment or increase the rate of profit by reducing real wages because it merely resulted in a fall in demand and costs and, therefore, in prices and investment. By the end of the Second World War the argument had been accepted widely enough to influence the authorities. Experience had already shown that reducing the money wage did not necessarily reduce the real wage. The index for money wages in Britain declined from 1920 until it reached its lowest point in 1934, though most of the decline, which Keynes described in 'A Tract on Monetary Reform', occurred in 1920-22.¹³² Yet, in both the early decline and during the depression, real wages rose. Keynes's argument applied to the closed economy and did not preclude the generation of employment in the open economy if lower money wages caused exports to increase, though it seemed improbable that that could offset the reduction of domestic demand, at least in the short term.

A second reason for assuming that money wages could not be reduced was that trades unions had grown stronger, especially in Britain, and had steadily made reducing money wages more difficult, all the more so by the unevenness across occupations of the growth of their strength and the unevenness in wage reductions that would have resulted. One effect has

¹³¹ Keynes, *The General Theory of Employment, Interest and Money*.

¹³² Routh, *Occupation and Pay in Great Britain 1906-1960*, Table 49.

been that Keynes's argument is sometimes misunderstood as being no more than the recognition that money wages and prices cannot be reduced.

For a new theory, therefore, firstly assumptions were needed that allowed the prices of each country's products to be raised or lowered in terms of the currencies of the other countries. Secondly, the influence of income or expenditure on demand had to be taken into account by modifying Mill's assumptions about demand.

In what seem to have been the first attempts at such a theory, Robinson¹³³ and Harrod¹³⁴ modified Mill's assumptions to make demand depend on income as well as prices. They also assumed that the authorities can influence investment and can, therefore, reduce unemployment by increasing investment, though the consequent rise in income results in more imports and a worsening of the trade balance. Devaluation without an increase in investment results in an improvement of the trade balance and reduces unemployment, though it does so by displacing production and generating unemployment in other countries. Such 'competitive devaluations' or 'beggar-my-neighbour' policies were held to have added to the economic distress between the two world wars. But a group of countries that trade mainly among themselves can reduce unemployment by coordinating to stimulate investment within each, thus causing income and employment to rise, while the consequent increase of imports of each is offset by the increase in imports of the others.

Harberger in 1950¹³⁵ and Alexander in 1952¹³⁶ seem to have been the first to take into account the identity of the excess of income over expenditure, which they term "hoarding", and the balance of trade. But they both assume that economic agents choose how much to hoard, rather than consider, as do Harrod and Robinson, saving and investment to be separate quantities, and that results in fully determinate systems in which nothing can change and the authorities have no functions.

Harberger uses mathematical equations to discuss several models with two countries, each producing one good. Only the models he calls "Keynesian" and "General" allow trade imbalances. In these models demand for the domestic and for the imported goods depends on income, measured as output, and on the exchange rate. The identity of the excess of income over expenditure and the trade balance is evident from the equations. Alexander's reasoning is entirely verbal. He starts with the identity and assumes that devaluation causes a reduction in real income and a

¹³³ Robinson, "Banking Policy and the Exchanges."

¹³⁴ Harrod, *International Economics*.

¹³⁵ Harberger, "Currency Depreciation, Income, and the Balance of Trade."

¹³⁶ Alexander, "Effects of a Devaluation on a Trade Balance."

proportional reduction in expenditure, or “absorption”, in addition to which he considers a “direct” effect on absorption from various causes, such as money balances, money illusion and redistribution of income. In discussing the last he mentions that devaluation can stimulate investment by the shift of income to profits it causes and can thus increase absorption, but he does not take up the point and he dismisses the direct effect as transitory.¹³⁷

To assume that economic agents decide how much to hoard is to assume that they decide what their countries’ trade balances should be, which implies that incomes and the exchange rate must be such as to yield consistency. It also means that consistency must be ensured between the balance of trade, a nominal quantity measured at current prices, and hoarding, which, if economic agents are rational, is a real quantity determined by nominal income and prices. Harberger assumes that ‘the marginal propensities to hoard, to import and to consume home goods operate not on national production *per se* but on “real consumer income,” of which national production is an adequate measure only when the exchange rate is constant’¹³⁸. Alexander’s assumptions are similar; he takes the fall in real income resulting from the effect of devaluation on the terms of trade as given and assumes that the effect on absorption is then determined.

Harberger’s assumptions are mutually consistent, but they determine all the quantities fully and do not permit discrete changes of the exchange rate. In each country income and the exchange rate, on the one hand, are related to imports and absorption of the domestically made goods, on the other, and to these relations are added the relations to hoarding. Then income, trade and the exchange rate are all determined and changing the exchange rate requires that one or more of the relations be changed. With Harberger this is concealed by his restriction of his argument to derivatives, i.e. infinitesimals. To get his formula for the derivative of the balance of trade with respect to the exchange rate he uses production as a measure of real income, on the grounds that it can be done “as long as the terms of trade do not change”¹³⁹. In other words, his formula is valid for notional, infinitesimal changes only. It is not the kind of approximation by which an elasticity is calculated using an arc elasticity in place of a derivative, for a normal devaluation in Harberger’s scheme would alter the terms of trade and, hence, the derivatives. Similar remarks apply to Alexander’s argument, though it is presented without the equations.

¹³⁷ Alexander, 273.

¹³⁸ Harberger, “Currency Depreciation, Income, and the Balance of Trade,” 52.

¹³⁹ Harberger, 52.

The Balance of Trade When Countries Coordinate: Meade

Another change from Mill's assumption is to assume that each country's authorities can determine the country's aggregate expenditure, comprising both consumption and investment. Meade seems to have been the first to do this. Where Robinson and Harrod had equated expenditure with income and Alexander and Harberger with income less private hoarding, Meade, in his 'Balance of Payments'¹⁴⁰, lets the authorities determine expenditure, which comprises both investment and consumption, though he refers to it as consumption.

As would be expected, Meade reaches a conclusion similar to that reached here, namely that countries can have balanced trade with full employment if the authorities adjust expenditure in both countries and the exchange rate suitably. But he reaches it by assumption that are unrealistic or unnecessary. First, he assumes that the authorities of both countries want trade to balance and thus ensures coordination. Then he follows the practice of assuming there are two countries, each producing one good that it consumes and exports to the other. Income is the wage cost of production. Each country's total demand is necessarily its expenditure and its demand for each good is a function of income and prices. To determine what happens when a country is at the limits of its production possibilities, which he terms full employment, Meade adds, first, that at full employment, an increase in demand results in a rise in money wages and, therefore, in inflation,¹⁴¹ and, second, that there is a level of demand in each country that gives full employment without inflation. No attention need be paid here to the distinction between Meade's two means of controlling aggregate expenditure, 'monetary policy' and 'fiscal policy', the former being the use of interest rates and banking measures, the latter consisting of changes in taxation and government spending, for the only differences in his exposition are the quickness of their effects and the effects on capital flows between countries, of which the former is irrelevant since only stationary states are considered and the latter because the concern here is the trade balance.

Meade's conclusion can be demonstrated by starting with trade not balanced and unemployment in both countries. By increasing expenditure the surplus country can increase imports and reduce exports, whilst the other country can reduce imports and increase exports by reducing expenditure. It is possible that the deficit country reaches full employment before trade is brought into balance, in which case it reduces expenditure further. If it is the surplus country that reaches full employment too early, increasing expenditure there causes inflation and, so, the deficit country must devalue

¹⁴⁰ Meade, *The Theory of International Economic Policy. Vol. 1, The Balance of Payments.*

¹⁴¹ Inflation here means rising prices. Meade uses the term for increased expenditure.

for an increase of its exports and a decrease of its imports, which, by themselves, create employment in that country and unemployment in the other and, therefore, require that it also increase expenditure to reduce the unemployment it has caused. This last step would also be required should the two countries reach full employment without inflation before trade is balanced. A succession of such steps can be taken to result in convergence to the intended outcome. In principle, the authorities of both countries can cooperate to reach the same outcome in one go.

Meade's assumptions, that countries specialise, that their costs of production vary only with wages and the scale of production and that countries agree on balancing trade, are both unrealistic and unnecessary. Countries do not specialise in this way and, if they produce competing goods, the costs of production cannot consist solely of wages; wage differences must be offset either by another cost, which is profit, or by protection. Neither do countries choose mutually consistent trade balances, something Meade's scheme cannot be used to discuss without adding assumptions to resolve inconsistencies. His assumptions are unnecessary since the conclusion he wants is similar to one reached here without them, namely that countries can coordinate to reach agreed mutually consistent trade balances by adjusting expenditure and that exchange rates need be changed only to allow those that are not at the limits of their production possibilities to reach these limits.

Not only are the assumptions unrealistic, but they also are inconsistent with Meade's own theory of trade. Having expounded the theory just described in 'The Balance of Payments', the first volume of his book, Meade uses the Haberler representation in the second volume, 'Trade and Welfare', to discuss the determinants of the pattern of trade, although the Haberler representation allows each country to produce more than one good in a range of proportions and several countries to produce the same goods over a range of prices. The assumption of specialisation is consistent with Ricardian theory, which would be a special case of Meade's theory of trade and inconsistent with the factor endowments theory. Yet this inconsistency is standard practice; their theories may vary, but textbooks that deal with both the balance of payments and the pattern of trade are alike in that they discuss them in separate sections with different assumptions.

Floating, Real, Shadow and Equilibrium Exchange Rates

Letting the Market Set the Exchange Rate

The preceding discussion of exchange rates has been concerned with the reasons that have been given as to why changing an exchange rate should have a specific effect on exports or imports. With minor qualifications, the arguments or theories that have been discussed here reach the conclusion

that devaluation improves the trade balance and revaluation worsens it. One consequence has been that the belief that the market should determine exchange rates has, over time, become the orthodoxy. Efforts by the authorities of a country to influence exchange rates are believed to be either misguided or, if they prevent a currency from appreciating, a form of unfair trade akin to export subsidies and non-tariff barriers against imports. Yet, in some countries the same authorities as believe in the orthodoxy have yet been trying to keep their exchange rates with their countries' most important trading partners stable or fixed. A second consequence, one of less practical importance, has been the creation of the notion of the real equilibrium exchange rate (RER or REER).

Friedman¹⁴² gave what are the standard arguments for allowing the market to determine the exchange rate. Other things the same, the demand for any country's currency is higher the lower its price in term of other currencies and the lower this price, the more favourable the trade balance. Hence, when a country runs a balance of trade deficit the supply of its currency increases and, other things the same, the price falls and that causes its trade balance to improve. As with the earlier theories of the gold standard, there is no reason for the authorities to interfere, in this case because the market knows better and leads the exchange rate to an equilibrium position.

At first sight the demand curve for the country's currency is like that for any good and is to be taken as given.¹⁴³ But economic theory derives demand curves for consumption goods from their intrinsic utility to their users and a foreign currency has no such intrinsic utility to its holder.

Friedman implies this when he gives the motives for wanting to hold a foreign currency. "Holders of foreign currencies want to exchange them for the currency of a particular country in order to purchase commodities produced in that country, or to purchase securities or other capital assets in that country, or to pay interest on or repay debts to that country, or to make gifts to citizens of that country, or simply to hold for one of these uses or for resale.... Other things the same, the more expensive a given currency, ... the less of that currency will in general be demanded for each of these purposes."¹⁴⁴

Now the demand for the currency is given by the demand for the country's goods, etc.; the cheaper the goods, the greater the demand for them and, hence, for the country's currency as the means of payment for them. Also, the residents want to spend less in terms of foreign currency because foreign goods

¹⁴² Friedman, "The Case for Flexible Exchange Rates."

¹⁴³ Friedman, 159.

¹⁴⁴ Friedman, 159.

are more expensive in terms of their currency. It is not reasoning that can be applied to securities and it must be reversed for debt service, but, leaving these caveats aside, it provides supply and demand curves of the desired kind for the currency. But the causation has been reversed and that eliminates Friedman's balancing mechanism; now the demand for the country's currency is equal to the payments that have to be made, in particular for the country's exports, and the reason that the currency depreciated in the first place, the lower price because of the greater supply, is missing.

The point may seem abstruse, but it indicates that demand curves for foreign currencies are not self-evident. Friedman is not explicit about the international payments system, but the motives he gives for wanting a foreign currency imply that all exports have to be paid for in the currency of the exporting countries, which is impossible. Importers must have the currencies to pay for imports. There have to be common means of payment.

Once a group of countries has such a common means of payment, for example gold, the US dollar or the IMF's Special Drawing Rights (SDRs), there is no reason why any country's exporters should expect or wish to be paid in their own currencies, or why their authorities should allow it. If several different means of payment are accepted by the group, it is to be supposed that they are easily exchanged for one another at little cost and, if expectations about exchange rates can be ignored, that exporters accept any of them with due allowance for eventual exchange costs and that their authorities do so too. Importers can use them indifferently as well, also with allowance for eventual costs of exchange. If a single means of payment predominates, exchange costs are limited to conversion between domestic currencies and that means of payment.

Even when there is a single preferred means of making international payments other currencies may be accepted and be used as reserves by monetary authorities. Such currencies are referred to as "hard". Currencies that are not hard are not normally accepted for international payments even by the exporters and authorities of their countries. An African, Brazilian or Indian exporter quotes his prices in dollars or some other hard currency and expects to be paid in that currency.

Up to this point Friedman's assertions about the effects of changes of exchange rate presuppose, explicitly or implicitly, that other things are the same, the usual *ceteris paribus* condition. But other things cannot be the same when the price of a currency changes; the exchange rate is a macroeconomic quantity. In contrast to the prices of goods, for which it is admissible to make the usual assumption, that a change in the price of one good has a negligible effect on the rest of the economy, a change of the exchange rate causes other prices, the purchasing power of monetary

aggregates and income distribution to change. Monetary aggregates also change when speculators buy and sell currencies. Friedman remarks, "Of course, after the event, the amount of a particular currency purchased must equal the amount sold – this is a question simply of double-entry bookkeeping."¹⁴⁵ Here he is mistaken, it is not bookkeeping but two ways of saying the same thing. Buyer and seller keep separate books. In the buyer's book the currency bought is entered as an increase in that type of asset and is balanced by decreases in other assets or increases in liabilities. In the seller's book the sale of the currency is entered as a decrease in that type of asset, with balancing increases in other assets or decreases in liabilities. Monetary aggregates change, normally because these balancing items are assets or liabilities of the banking system. This is just the common effect of the trade balance on the money supply; a surplus increases it and a deficit decreases it, unless there are offsetting capital and factor payments or actions by the monetary authorities. Friedman's remark diverts attention from this by making it seem that the purchase and sale of the currency are what they cannot be, balancing items in the same book.

Friedman's demand curves cannot be supposed to exist. But he has another argument for allowing the market to determine the exchange rate, one that does not follow from the demand for the country's currency. Friedman assumes that there is an equilibrium price for the currency and that speculators will find it. Much of his essay is devoted to this argument.

An equilibrium is assumed, but must first be defined and Friedman is neither consistent nor clear in this. Initially the argument is confined to the effects of changes of exchange rates on the trade balance, although that is just one component of the balance of payments. But, later in the essay, he switches to discussing the balance of payments, rather than the balance of trade, stating that a currency tends to appreciate because of an incipient surplus of receipts over payments, i.e. excess demand for the currency, and tends to fall in the reverse case.¹⁴⁶ As already mentioned, he specifically allows purchases of securities and debt service to be reasons for wanting a particular currency, and, when he refers to surpluses and deficits of receipts over payments, he appears to allow for capital movements and items like remittances from workers abroad and from emigrants, on which several low wage countries depend.

Equilibrium, then, concerns the future as well, for debt has to be serviced. Friedman asserts that there is a "final position" of the exchange rate and asserts that fundamental factors, if they are generally regarded as likely to be permanent, can produce a change of an exchange rate and that speculative

¹⁴⁵ Friedman, 160.

¹⁴⁶ Friedman, 161.

transactions "... hasten its approach to its final position".¹⁴⁷ If it is flexible and currency markets are free, the exchange rate adjusts to make the amounts of the currency demanded and offered equal all the time, though the ultimate adjustment to "... a change in external circumstances ..." takes time, anything from months to a generation.¹⁴⁸ It seems from this that the final position is something inherent to the real economy. But, if, for instance, there were at some moment to be a change of such external circumstances or fundamental factors as require a generation to adjust to and then no further change, is it plausible that the final position would be independent of monetary and other policies in the interim? Or are these, too, fundamental factors? Beyond the clearing of the market for the currency at every moment, Friedman's equilibrium is an ill-defined assumption.

Here, too, there is an unwarranted assumption of *ceteris paribus*, this time for other countries. If the trade balance can affect the money supply, the monetary policies of other countries can affect their trade balances and, therefore, that of the country in question. It is true both of the short run and of the final position. Friedman does not explicitly say so, but it can be seen from an example he gives; depreciation prevents monetary deflations in other countries from forcing down the prices of foreign goods in terms of the domestic currency and it is also caused by domestic inflation.¹⁴⁹ Each country's exchange rate is, therefore, affected by the monetary positions of other countries and cannot reach its final position until all countries have appropriate monetary positions. Countries must, therefore, coordinate their monetary policies if their exchange rates are to reach their final positions; speculative transactions are not enough.

Others, apart from Friedman, have given arguments in favour of letting the exchange rate be determined by the market. Their arguments need not be discussed individually; they differ only in the details and it can be shown that their common assertion, namely that the markets will bring currencies to their correct exchange rates if left to do so, cannot be true.

For it to be true, the foreign currency markets would have to agree on the relation between the exchange rate and the trade balance for each country. Virtually all those who take part in the foreign currency markets agree there is some such relation, but there is no explicit agreement as to what it is. Financial institutions and other firms often have their various models and formulae, which differ and are sometimes proprietary secrets. There is no consensus, as can be seen from, among other things, the lack of a generally accepted formula for the RER, and, if there were to be one it would

¹⁴⁷ Friedman, 162.

¹⁴⁸ Friedman, 182.

¹⁴⁹ Friedman, 182.

not be objective, but by agreement, because there is no such relation. Any country's trade balance depends on the expenditure of other countries and both trade and current account balances of all countries must add to zero. If markets were to allow for this adequately, they would have to have reliable up to date information on production, saving and investment for each country at the moment and the immediately preceding weeks and to forecast them for the near future, whereas it takes months to have such information in the most advanced countries. Both the present and the future are forecasts, which, though useful, are unreliable.

In practice, those who take part in or try to influence the foreign currency markets, especially the central banks and other financial institutions, are inconsistent in their belief that floating exchange rates are rationally determined. On the one hand, forward rates, i.e. the contractual rate now for exchanging currencies at a specified moment in the future, are rational in the sense that no profit can be made by buying a currency and selling it forward while receiving interest on it in the meantime. A gain or loss on the difference between the interest rates of the two countries is offset by the difference between the spot rate and forward exchange rate, what is termed covered interest parity. If a country has a higher interest rate for that period, its currency depreciates on that moment's forward rate, as opposed to the rate that actually occurs later. Arbitrage brings this about and empirical tests show that it does so reliably and accurately under almost all circumstances. On the other hand, economic commentators and central banks believe that raising the interest rate causes the currency to appreciate. Central banks regularly try to prevent depreciation of their currencies by raising interest rates or to prevent appreciation by lowering them. A higher three month interest rate results in the currency depreciating on its three month forward rate, though the authorities who try to influence the exchange rate expect the opposite for the actual rate, that capital will be attracted and bid up the price of the currency. The actual exchange rates three months later are not what the forward rates were, except by chance, which is to say that uncovered interest parity does not hold. Empirical studies have verified that it does not hold as a rule and markets do not believe it holds.

The Equilibrium Exchange Rate (RER, REER, EER)

Friedman assumed determinacy in the long run by positing a final position for the exchange rate determined by fundamentals and, therefore, something inherent to the economy. Krugman agrees with Friedman that there is such a final position and states the problem of finding it when he says that '...the equilibrium *real* exchange rate at some time in the future will be foreseeably different from today's real exchange rate ...', and '... that policy toward the *nominal* exchange rate can somehow facilitate the adjustment

toward this future real exchange rate.¹⁵⁰ Nominal prices are not fundamentals and, as Krugman points out, the exchange rate at which the market clears is not necessarily the exchange rate of the final position. So the nominal exchange rate of the moment may have to be corrected by the appropriate policy, which means that there is a correct exchange rate at any time. In particular, since Krugman has a real exchange rate (RER), the nominal exchange rate must, at least, be corrected for the price level. The final position is a real equilibrium exchange rate (REER).

Judging from Krugman's statements, it seems that the REER cannot be easily calculated by economists or government officials. It is, therefore, best left to be discovered by the market, which was Friedman's opinion. If the final position were just balance on the trade account, calculating it from the assumption that the trade balance is determined by the exchange rate would, in theory, be simple. But Friedman and Krugman agree that the final position or REER is not necessarily one at which trade balances. Ideally markets take into account the other relevant components of the balance of payments; among other things, capital movements, the consequent debt and debt service, as well as the yields on foreign investment, the pay of non-residents and the remittances of residents and non-residents. For instance, a country that has accumulated debt not denominated in its own currency may have to run trade surpluses to service that debt.

If there were a valid theory it would be possible, in principle, to determine by deduction the fundamentals and the formulae relating them to the trade balance and the exchange rate and, because the theory supposes determinacy, the only diversity there can be is in the estimates of the various quantities involved. In practice, there is no standard list of fundamentals and there are no theoretical formulae for the relation. They are matters of personal judgement and statistical correlations with correspondingly many different equilibrium exchange rates (EERs).

A brief account of some of the EERs and how they were derived suffices to illustrate the shortcomings of such notions. Among the quantities that have been used as fundamentals are differences in interest rates between countries, the share of GDP consumed by the government, the terms of trade, the relative prices of tradables and untradables, the ratio of output per head in the production of tradables to that of untradables (as an indicator of the difference in technical progress between the two sectors), the old age dependency ratio and the stage of development (represented by GDP per head). Each EER has its own list of fundamentals and is estimated statistically using linear relations. Among the EERs so derived are FEER, DEER, BEER, ERER (Consultative Group on Exchange Rates) and NATREX.

¹⁵⁰ Krugman, "Equilibrium Exchange Rates," 160. Italics in the original.

EERs are not meant to yield actual exchange rates, but exchange rates that should be. Differences between actual nominal exchange rates and those obtained from the EER are termed “misalignments”. If the currency is priced above the EER, the misalignment is termed “overvaluation” and if it is priced below, “undervaluation”. Misalignments of the exchange rates of high wage countries can be expected to be corrected by the market, but, in countries that do not have reliable markets for foreign exchange, notably the low wage countries, it is the authorities who must correct them. These definitions of the terms “misalignment”, “overvaluation” and “undervaluation” are specific to this context. In other contexts the terms may be defined differently, or may be left undefined.

Different EERs give different measures of misalignment, so an institution like the IMF, of which the advice to low wage countries on exchange rates has authority, must select one procedure for calculating EERs and make that the standard. The IMF does not necessarily have to give different countries mutually consistent advice on all matters; it advises each country according to what it judges to be right for that country and, if its advice to some countries is detrimental to others, it is not the duty of the IMF to arbitrate. But EERs have to be mutually consistent because they are supposed to give equilibrium. For these reasons the IMF set up the Consultative Group on Exchange Rates (CGER), which has adopted the methods proposed by the staff of the IMF for selecting fundamentals and estimating EERs as expounded in ‘Methodology for CGER Exchange Rate Assessments’¹⁵¹.

A description of the IMF’s way of obtaining its EERs illustrates the arbitrariness of such exchange rates. Three different methods are proposed. For each of the first two methods a list of fundamentals is chosen and a cross country statistical estimate is made of the relation between the fundamentals and the current account for the first method and the RER for the second. In the first method the relation is used to calculate a current account “norm” for each country from its projected medium term values of the fundamentals, and the change in the exchange rate needed to reach it, calculated by using elasticities of imports and exports with respect to the exchange rate, gives the RER for that norm, which is the EER. In the second method the EER is the RER obtained from the medium term projections of the fundamentals and their relation with the RER. As expounded, the two methods do not use the same lists of fundamentals and the number of countries used for the estimations are 54 in the first method and 11 in the second. The third method is to decide on a value of net foreign assets for each country and, as with the first method, to use elasticities, to calculate the necessary change in the exchange rate.

¹⁵¹ IMF. November 8, 2006.

Standardisation in this way has the drawback of ignoring the specific characteristics of individual countries that caused those countries to diverge from the assumed relations. It would, perhaps, be justified if it were certain that the choice of fundamentals was so comprehensive that the characteristics omitted were of little importance. But this is explicitly not the case; '... by including country specific constants, the resulting estimates of country effects may be unduly influenced by historical realizations ...'¹⁵². In the exposition of the first of the three methods, it is stated, regarding some variables, '...that the impact of these variables on the current account may be weakened by country-specific factors ...' and 'population growth has a stronger economic and statistical effect across countries than over time ...'¹⁵³. Moreover, 'Other variables whose economic and statistical significance is mostly captured by country-specific constants ... are excluded from the regression'¹⁵⁴.

There is no reason why the exchange rates for different countries obtained by any of these methods should be mutually consistent. So a procedure, which need not be described here, has been devised for modifying the exchange rates to make them consistent, though no explanation is given as to why the new ones should be considered as good as the original ones. There is also no discussion of how to make the trade and current account balances mutually consistent.

Apart from the arbitrariness of the choice of fundamentals, this and other methods of estimating the EER are arbitrary to the extent that the trade balance is. Unless the country's trade balance is given by constraints on its financing possibilities, any procedure for fixing the trade balance, such as deciding on the net financial assets at some time, is necessarily arbitrary. A fallacious procedure for making the trade balance determinate that has been used in several methods of estimating EERs is to assume that saving and investment are determined by the characteristics of the country, e.g. propensities to save and rates of return on investment. Then the identity of the excess of saving over investment and the trade balance yields the trade balance and, since the trade balance is also assumed to be determined by some form of exchange rate, usually the RER, it yields the exchange rate. The fallacy is that an identity holds always and is not a condition that has to be met. If the exchange rate determines the trade balance it determines the difference between saving and investment, in which case saving and investment cannot be independently determined by propensities to save and returns on investment. Alternatively, if saving and investment are determined by the characteristics of the country, so is the trade balance and the exchange rate is irrelevant.

¹⁵² 'Methodology for CGER Exchange Rate Assessments.' p.8.

¹⁵³ *Ibid.* p.10

¹⁵⁴ *Ibid.* p.11

Procedures like these are self-contradictory; they limit the effects of the exchange rate to the trade balance, on the one hand, but are, on the other hand, justified by the claim that the efficiency of resource allocation depends on 'getting the exchange rate right', a claim that presupposes that the exchange rate's effects on the economy are pervasive. In other words, either the fundamentals are unaffected by the exchange rate or some must be recalculated for each exchange rate. Sometimes the EER is calculated for the present, but usually it is calculated for the medium or long term, with the implicit assumption that it will hold indefinitely, barring unforeseen events. Since the fundamentals are projected first and the EER is derived from them, the RER is being assumed to have no influence on them or on the allocation of resources; the effects of the exchange rate on the economy are ignored.

Shadow Exchange Rates

The self-contradiction can be seen from the methods that have been proposed for assessing existing industries and proposed projects in low wage countries that start from the assumption that free trade gives an optimal allocation of resources. These methods and the controversies they provoked may be out of date, but they explain the evolution of some economies because they were used to determine policy in low wage countries and, hence, illustrate the type of theory involved. They all use orthodox theory and represent trade in the manner of Haberler and Meade. Since there is a consistent set of world prices at which all tradable goods can be exchanged, the most economically efficient pattern of production and trade is that at which domestic and world prices are equal. Free trade is also assumed to result in an equilibrium exchange rate.

Before the 1990s, especially during the 1960s and 1970s, low wage countries commonly protected their industries against foreign competition, often taxed or subsidised their exports and kept their exchange rates fixed by controlling foreign exchange transactions. Domestic prices were not equal to world prices and the currency could, according to orthodox theory, be presumed to be overvalued since domestic production was protected and controls were needed to prevent the depletion of the foreign exchange reserves or depreciation of the currency. Investment and production using those prices and that exchange rate were presumed to be inefficient; hence the need for methods by which existing activities and future projects could be assessed.

It follows from the assumptions about free trade that the economic gain from an activity is given by the foreign exchange it adds or saves, which is the value added of that activity calculated using world prices. If an activity is not competitive at world prices it must be protected against foreign competition and its value added at domestic prices is greater than at world

prices. Balassa¹⁵⁵ and others, therefore, proposed comparing the two added values and called the ratio of the former to the latter the “effective rate of protection” (ERP). In principle, therefore, the domestic price of a good less the cost at domestic prices of the inputs needed to make it is compared to the world price of the good less the cost of the same inputs valued at world prices. An activity with an ERP greater than one is being protected.

A good cannot be produced at less than the world price, given the prices of the inputs, because international competition results in the lowest prices. Hence, no manufacturing activity in a low wage country can have an ERP less than one, subsidies apart. An activity’s value added can be negative at world prices and yet be positive at domestic prices because of protection. Protection allows the price of the output to be raised relative to the world price by more than the prices of the inputs, usually because the tariff on the output is greater than the tariffs on the inputs. An activity with a negative value added at world prices causes the economy a foreign exchange loss and ought to be stopped.

Even if these arguments are accepted, the ERP cannot, by itself, be a guide to the economic efficiency of an activity because it omits the costs of the values added. As formulated here, it only take into account the costs of current tradable inputs, like cotton in spinning or coal and ore in making iron, which are physical quantities with world prices. But it does not explain what are the costs in a period of capital equipment, infrastructure, etc., which are used for several periods. Then there are untradable current inputs, such as energy and local transport, which do not have world prices. Assuming some method for ascribing suitable prices to the untradable current inputs, the costs of all current inputs are known. Deducting these costs from the value of the output, which is assumed to have a world price, leaves the value to be ascribed to the inputs of capital, infrastructure, etc.. There are, therefore, two sets of quantities that have to be determined, the prices of untraded goods and the rates of return on fixed assets, both tradable and untradable.

If these quantities are not known the only criterion that follows from the method is whether or not the ERP is greater than one; any activity for which it is above one causes an economic loss and should be stopped. Its proponents usually hesitated to go so far. They would find that the ERPs for various countries were high and claimed that this showed economic inefficiency, but they ranked activities and projects by their ERPs and rejected those that were too high according to some criterion. An example is that of Bacha and Taylor, who recommended calculating the exchange rate at which the ERP showed the activity to be competitive at world prices and accepting the activity if the devaluation implied was not greater than that needed to

¹⁵⁵ Balassa, “Tariff Protection in Industrial Countries: An Evaluation.”

reach the free trade exchange rate.¹⁵⁶ They proposed calculating the free trade exchange rate by using price elasticities of demand on the assumption that income and monetary quantities had no influence. They did say, 'Completeness would require the explicit incorporation of monetary variables ...', but added, '... it is simpler to keep money in the background ...'.¹⁵⁷ They assumed without demonstration, as is usual in such exercises, that free trade results in a determinate exchange rate that is also optimal in some way. There is no further need here to refute that. The assumptions of Bacha and Taylor illustrate the point; monetary variables cannot be ignored, nor can the effects of changes of exchange rates on incomes through the effects of real wages and the rates of return on investment, and on the prices of untradables.

Numerous estimates of ERPs of goods produced in low wage countries were made in the 1960s and 1970s and few were not greater than one. These were mostly simple exercises because it was assumed that world prices could be calculated from domestic prices by deducting the tariffs on imports, giving a formula for the ERP consisting solely of tariff rates and coefficients of units of inputs per unit of output. It was easy to conclude from the results that the economic difficulties of low wage countries were largely explained by the inefficiency of their investments, especially those with negative values added at world prices. All a country had to do, therefore, was remove trade barriers and accept the more efficient allocation of resources of the free market, including the ideal exchange rate. Some allowance could be made for the infant industry argument by allowing moderate protection, but should not be allowed to cause a misalignment of the exchange rate.

Since the ERP procedure follows from *a priori* reasoning and several assumptions, it should have been empirically tested, especially because of the consequences when applied, as it often was, to low wage countries. In the natural sciences this would have been a matter of course. One test would have been to compare between countries the supposed border prices obtained by the ERP method for individual goods, both outputs and inputs. It would have been a test of the reliability of the method and would have shown that these prices differed from country to country. During the period of the popularity of ERPs and DRCs the high wage countries produced many of the manufactures that the low wage countries produced and protected their production, and that should have refuted the notion of world prices. Balassa, himself, provided evidence in a study published in 1965 of the effective protection of the high wage countries, which showed that these countries protected their industries, in particular, those that competed with

¹⁵⁶ Bacha and Taylor, "Foreign Exchange Shadow Prices: A Critical Review of Current Theories," 222.

¹⁵⁷ Bacha and Taylor, sec. IV.

imports from low wage countries. A second test would have been to use the prices of the high wage countries as world prices to calculate the ERPs of the low wage countries. If the result was less than one the explanation of the higher results that had been obtained for low wage countries could not be inefficiency. Balassa's study gave evidence for that by showing how much protection the high wage countries thought they needed; it showed that the protection of the high wage countries was progressive, i.e. the tariff on an import competing with a final output was higher than on the imported inputs, and gave more protection than a uniform tariff. Thus the tariff on imported yarn was higher than on raw cotton and on cloth higher than on yarn. A third test would have been the converse of the second, to calculate the ERPs of the high wage countries using the domestic prices of the low wage countries in the numerator. It would be expected that, because of the low wages, the value added of the low wage countries would be low and the calculated ERP less than one. Though simple and obvious, none of these tests seem to have been contemplated.

Bruno and Krueger proposed a method of assessing activities and projects that, on the same assumption of world prices as Balassa, purports to take account of costs.¹⁵⁸ In their method activities are to be compared according to their costs in terms of domestic resources (DRC) for earning or saving a unit of foreign exchange. It depends on how factors, including foreign exchange, are priced. If actual factor prices are chosen, corrected if need be for imperfections of factor markets, the cost of a unit of foreign exchange is its opportunity cost in terms of actual domestic prices. Balassa and Schydlosky objected that the DRC did not give the opportunity cost in terms of foreign exchange because the domestic prices of goods and, hence, of factors were determined by tariffs.¹⁵⁹ Bruno's solution was to use shadow prices determined as the outcome of a planning exercise. He described production as a linear input-output system with goods and factors, the factors including labour, capital and foreign exchange, and the planning exercise was to maximise a linear function of some goods with constraints on the supply of factors.¹⁶⁰ His shadow prices were the duals of the constraints, and an activity that produced more than it cost in terms of these shadow prices yielded a net benefit.

A general objection to such shadow prices is that they do not apply to an actual economy. The dual to a constraint is the ratio of the increase or decrease of the maximand to the increase or decrease of that constraint that gives that change of the maximand, provided the binding constraints remain

¹⁵⁸ Bruno, "Planning the External Sector"; Krueger, "Evaluating Restrictionist Trade Regimes: Theory and Measurement."

¹⁵⁹ Balassa and Schydlosky, "Domestic Resource Costs and Effective Protection Once Again."

¹⁶⁰ United Nations. and Bruno, *Planning the External Sector*.

the same. It is irrelevant to an economy in which the program is not carried out, and planners cannot make households consume and save or firms produce and invest according to their maximand. A maximand is not an extraneously given objective formula; it is chosen by the planners and, being arbitrary in that sense, its associated shadow prices are too. It is, for instance, possible for the net benefit of an activity to be positive for one choice of maximand and negative for another.

A second objection is to suppose foreign exchange is a factor. As a factor it must have a shadow price and, therefore, a constraint, which has to be the limit to the trade deficit that Bruno assumes. If any of the goods in the maximand are tradable, this constraint is binding and its dual is positive. Bruno interprets the dual as the shadow price of foreign exchange, which, to economists accustomed to thinking of dual variables as shadow prices, seems the only price the dual can be.

This is fallacious reasoning, for an exchange rate relates two currencies and, while world prices are quoted in terms of a foreign currency, no mention has been made of a domestic currency. In this case the dual variable is the ratio of an increase of the maximand to an increase of the trade deficit, as already explained, and is not a price since the deficit is neither a commodity nor a currency. Doubling the weights of the maximand obviously doubles the dual variable although nothing else changes. Bruno gives the appearance of an exchange rate by weighting the goods in the maximand solely by domestic prices, though that makes the exercise self-contradictory by creating two sets of prices for tradables, those in the maximand and the shadow prices. No explanation is given as to how they compare. If, for instance, all the goods in the maximand are tradable and their relative prices are the same as their relative world prices, an exchange rate is being assumed, and, if the relative prices differ, the program is inconsistent with the principle that world prices should be used.

A third objections is that, like the ERP of Balassa, the DRC does not provide a comparison of the economic gain from an investment with its cost, despite the deduction of the cost of capital valued at a shadow price. Like the ERP, it cannot rank activities according to their economic benefits. Beyond requiring that the net benefit be positive after deducting the costs of capital and other factors valued at shadow prices, the DRC does not distinguish between investments. Leaving aside considerations of income distribution, environment and various social concerns, if an activity is to be evaluated as an investment, its economic benefit must be taken as a proportion to the investment. If, as is common for low wage countries, the benefit is the income generated by the investment, the criterion is the ratio of value added to the investment, usually the output:capital ratio. Neither

the ERP nor the DRC uses a ratio of this sort. Comparing different investments by their net benefits per unit of output is meaningless since it depends on the units in which the output is measured.

Evaluating Investments

Accepting that the supply of foreign exchange is the main constraint on investment in a low wage country, the three principal objections to the ERP and the DRC are that they do not compare the economic gain from an investment to the cost of the investment, that there are no world prices for most goods and that shadow prices obtained from optimisation exercises are irrelevant. All three can be avoided. If an industry or project is intended to earn or save foreign exchange, the foreign exchange gains and costs can be the unit prices of exports and imports as given from the country's trade data. This is sometimes more laborious than consulting tariff manuals, but is unavoidable. The logical way of comparing the benefit to the cost is the ratio of the benefit to the cost, for instance the ordinary output: capital ratio of development economics. For that the benefit each period, which is a flow, has to be compared to the cost of the investment, which is a stock. Prices have to be found for the untradable current inputs and for the tradable and untradable fixed capital, and a rate of return for the investment. Since these exercises use input-output tables, it is necessary to follow Schwartz and have two tables, one for the current inputs and one for the capital stock.¹⁶¹

Then, for each activity, a set of prices of untradable goods and a rate of return on fixed capital can be derived from the border prices on the assumption that the rate of return is uniform throughout. This is determinate, as can be seen from the formulae below.

There are n goods, the first u being untradable and the remainder tradable. A unit of good j requires a_{ij} units of good i as current input, b_{ij} units as capital good and l_j units of labour. The nominal price of labour is w and the price of good i is p_i , where the prices p_{u+1}, \dots, p_n are the appropriate border prices. Putting the rate of return as ρ , for each good produced in the country:

$$p_k = \sum^n p_i \cdot a_{ik} + \rho \cdot \sum^n p_i \cdot b_{ik} + w \cdot l_k$$

$$p_k = \sum^u p_i \cdot (a_{ik} + \rho \cdot b_{ik}) + \sum^{n-u+1} p_i \cdot (a_{ik} + \rho \cdot b_{ik}) + w \cdot l_k$$

For each untradable, k :

¹⁶¹ Schwartz, *Lectures on the Mathematical Method in Analytical Economics*.

$$\sum^u p_i \cdot [\delta_{ik} - (a_{ik} + \rho \cdot b_{ik})] = w \cdot l_k + \sum^{n_{u+1}} p_h \cdot (a_{hk} + \rho \cdot b_{hk}).$$

$$\delta_{ik} = 1 \text{ if } i=k, 0 \text{ if } i \neq k.$$

$$p_k = [w \cdot l_k + \sum^{n_{u+1}} p_h \cdot (a_{hk} + \rho \cdot b_{hk})] \cdot [\delta_{ik} - (a_{ik} + \rho \cdot b_{ik})]^{-1}.$$

$$k = 1, \dots, u.$$

These equations determine the rate of return as an implicit variable. Then, for each tradable good, the first equation allows the rate of return to be calculated from the prices of other tradables, untradables and the wage. The prices of untradables are obtained from the last equation. Each tradable good has its own rate of return and the criterion for choice is to prefer the goods with the higher rates.

The assumptions can be modified in several ways without change of procedure. Depreciation can be allowed for; one way is to increase the coefficients of each type of capital good in a fixed proportion, for instance by a small amount for buildings and by larger amounts for machinery, and another is to assume that each type of capital good lasts a certain number of periods, so that each is depreciated by an amount depending on its age and the rate of return. Underutilised capital stock in any given sector can be allowed for by modifying the assumption that additional production requires additional investment in that sector. Workers can be split into different categories with different wage rates or planners can attribute a shadow price to each category of labour. Alternatively, the wage can be set at zero, which is equivalent to answering the question as to what investment gives the most income, i.e. a form of output:capital ratio.

3. INTERNATIONAL PAYMENTS: THE BRETTON WOODS SYSTEM AND AFTER

There could be no presumption at the end of the Second World War, as there had been at the end of the First, of reverting to a system of international payments that had existed before. Returning to the classical gold standard was neither feasible nor wanted and what had followed it had failed. Now, about seventy per cent of the world's monetary gold was in the US, where the price of gold was fixed at \$35 an ounce, and little was left in Western Europe. Were the countries of Europe to adopt the classical gold standard, they would, under the best of circumstances, have required loans from the US as reserves just to maintain payments, as had the UK in returning to the gold standard seven years after the First World War, and they would have had to restrain expenditure despite the destruction and poverty. This the experience of the aftermath of the First World War ruled out, besides which it was unlikely that any restraint could have prevented capital flight and the prompt loss of all reserves. For the

US it would have been a choice between providing aid to alleviate the hardships and help recovery or providing loans for balance of payments purposes that would probably end in the US as flight capital.

A new system had to be devised that would avoid the misfortunes of the recent past without there being a precedent as guide. Only the US and the UK were in a position before the War ended to discuss what the new system should be. Their discussion began well before the end of the War and ended as an international agreement of almost all countries or their imperial powers at Bretton Woods, New Hampshire, in the US in July 1944. Since the US and the UK were the main industrial countries of the West at the end of the War, what they had negotiated could not be much influenced by the other countries when it came to signing the agreements.

The Requirements of a New System

Some of the requirements of a new system were clear from the start. One was an internationally accepted means of payment, at least a means of settling obligations between countries, and another was that exchange rates be fixed. A third was that the system be multilateral, that all participating countries be able to settle their international obligations through it. Fourthly, a requirement of the US was that the payments system be separate from the financing of Europe's reconstruction, which was to say that whatever its financial assets, they should not be used for financing persistent trade deficits.

In the immediate aftermath of the war the first requirement was met by the dollar, which, along with gold, was the only generally accepted means for settling international transactions. Sterling was officially a reserve currency like the dollar, but the UK '... did not create international liquidity on a noteworthy scale'.¹⁶² But, because of the fourth requirement, the dollar became for several years the only generally accepted currency for settling international transactions. The sole alternative to the dollar would have been a new currency or unit not backed by gold or dollars for settling accounts between countries, either of which could have been created simply by international agreement. That might have provided the necessary liquidity for countries to reduce controls and open their current accounts, but, under the circumstances, it would have resulted in the full amount created, or nearly, being used to pay for US exports, i.e. the US would have been making interest free loans of indefinite duration to an international institution. If it were to be a large amount, it would finance some of Europe's reconstruction, which the US, having cancelled \$20 billion of Lend-Lease debt, did not at that stage want to do in this way. If the amount were to be small, it would just be a superfluous adjunct to the dollar and gold. Hence, the main

¹⁶² International Monetary Fund., *International Reserves: Needs and Availability.*, 327.

proposal for a new means of payment, Keynes's International Clearing Union, was out of the question. It was to have been an overdraft facility accorded to the central bank of each member country and limited to a specified amount of a new means of settling international obligations, *bancor*, for a total of \$ 26 billion. If it had been adopted the US would eventually have acquired the \$23 billion worth of *bancor* of the other countries in payment for exports.

Both American and British officials wanted the second requirement, fixed exchange rates, to prevent the competitive devaluation that they believed had been one of the causes of the disorder of the inter-war years. Since a new international monetary system was being designed practically from scratch, an institution embodying the agreements had to be created. This was the International Monetary Fund, of which the countries were members bound by the Articles of Agreement. Most currencies were fixed in terms of the US dollar, which was the natural standard, and a few in terms of gold without being, themselves, directly on gold standards. Some of these currencies had currency zones of which they were the standard against which the other currencies of the zones were fixed. In place of the classical gold standard, there was a sophisticated gold standard; the currency issue of the US was limited by the amount of monetary gold priced at \$35 the ounce, though members of the public could not exchange dollars for gold at the US central bank, and the central banks of countries member of the IMF could exchange dollars and gold among themselves and to use their holdings of these as reserves for the issue of their own currencies. Exchange rates could be changed, but only as a last resort. It was an arrangement expected to allow adequate growth of economic expenditure in the US and Europe and of international reserves because the US stock of monetary gold was so much in excess of the limits of the currency issue that it was not expected to be a constraint on the money supply for the foreseeable future.

All outside the Soviet bloc agreed in principle on the third requirement, that the payments system be multilateral, but it did not mean that there could not be other payments zones at the same time. A payments system in which all participating countries could settle in dollars did not prevent payments zones, in each of which the currencies could be converted into the zone's standard currency within limits, but not into gold or dollars. Countries with empires constituted zones of the latter kind, the most important being the Sterling Area associated with the British Commonwealth. But Belgium, France, Holland, Portugal and Spain also had empires and associated currency zones. There could also be altogether separate payments systems, an example of the latter being the Soviet bloc. (Hence, here "Europe" refers only to countries outside the Soviet bloc.)

Having more than one payments system necessarily involved discrimination between countries belonging to any particular system and the others, and was, therefore, opposed by the US, which insisted on non-discrimination. But non-discrimination required in practice that the current account be convertible, for there would otherwise be controls over imports and there could be no checking that the authorities did not discriminate when they decided on individual transactions. It had, therefore, to be postponed. At first, when the European countries resumed trade with virtually no gold or dollars, it was necessarily as bilateral exchanges, little more than barter. There was no clearing arrangement yet for multilateral payments and, when several countries joined in a bilateral clearing arrangement, the Intra-European Payments Agreement, in June 1948, needs were still too urgent to accumulate reserves through current trade surpluses. So, countries avoided as far as possible giving credit, i.e. having trade surpluses, whilst reserves were too precious for them to use the little they had for making payments, i.e. to have trade deficits.¹⁶³

European countries with empires also used the associated payments or currency zones to alleviate some of their shortages of goods, though, in the case of the largest, the sterling area of the British Commonwealth, this was limited by the autonomy of some of the members and the objection of the US to discriminatory arrangements. Australia, India, New Zealand and South Africa were all financially and fiscally autonomous and had, through their trade surpluses with the UK during the war, accumulated sterling credits that were too big to be eliminated through trade in the foreseeable future. To meet the objections of the US to the sterling area, although there was no prospect of it gaining the necessary trade surpluses, the UK agreed to make sterling convertible in 1947 with the backing of loans totalling \$5 billion from the US and Canada. Convertibility was also considered a prerequisite for sterling to become a reserve currency, one that could be used for international reserves along with the dollar, as it had been before the war. Less concern was shown over the currency zones of Belgium, France, Holland, Portugal and Spain, none of which allowed its constituent parts any autonomy.

Institutions and Mechanisms

In addition to these requirements, there had to be mechanisms for avoiding balance of payments crises, the inability to meet payment obligations in foreign currencies because of a lack of reserves, i.e. generally accepted means for making international payments. As administrative control over trade and foreign exchange were removed and the current account became convertible, the only mechanism was for those with deficits

¹⁶³ Wilcox and United Nations Conference on Trade and Employment, *A Charter for World Trade*, 12–13; Tew, *The Evolution of the International Monetary System: 1945-81*.

to reduce expenditure to improve their trade balances and for those with surpluses to increase it to worsen theirs. All countries except the US controlled their capital accounts, so the primary reason for a loss or gain of reserves could be assumed to be the trade balance. Since the authorities of a deficit country might not succeed in bringing about the needed improvement in time, a source of financing was needed to supplement the country's reserves for the time needed. White, the leader of the US delegation in the negotiations, proposed a stabilisation fund, to which the members were to contribute their own currencies, gold and dollars according to the sizes of their economies and which would provide medium term loans. That fund was the basis for the final agreement and became the International Monetary Fund, which, with total contributions of \$8 billion, administered the system of fixed exchange rates.

As envisaged then, the purpose of the IMF was to keep exchange rates constant; any change was to be approved by the IMF. But the possibility had also to be allowed for that a country might have perpetual balance of payments problems or high levels of unemployment at its current exchange rate. An exception was, therefore, made for when the IMF judged that an economy was in "fundamental disequilibrium", for which no criterion was ever specified but which allowed devaluation by an agreed amount. In practice the criterion became the frequent recurrence of trade deficits despite the country having taken all the appropriate measures. If the IMF was to provide loans to a country that was running out of reserves because of trade deficits, it had normally to be satisfied the measures proposed by the country's authorities were adequate, and in time that led to a standardised set of procedures with indicators of performance. If trade deficits persisted nevertheless, devaluation seemed the only recourse. Thus came about a compromise between the US's preference for fixed exchange rates, logical for a country on the gold standard, and Keynes's orthodox belief, which he never gave up, in the efficacy of exchange rate changes, the outcome being a hybrid system, part gold standard, part fixed but changeable exchange rates, a system of which the two decades and more of its functioning were the time in which the economies of the US, Western Europe and Japan performed better than they had done before or have done since.

There was no compromise over allowing the payments system to finance reconstruction; financing of reconstruction by the IMF was explicitly prohibited. For that the US proposed the International Bank for Reconstruction and Development (IBRD), also known as the World Bank, which would sell bonds on the financial markets, which were at first in the US, and make loans from the proceeds. Its bonds outstanding were limited by its nominal capital, initially \$10 billion, and were, therefore, guaranteed, although only a small part of the capital needed to be paid up.

Europe: the Danger of Depression and Economic Recovery

If, as was believed, the classical gold standard had been too restrictive and to have caused unnecessary depressions, the Bretton Woods system might have been even worse, despite having been put together by people who wanted to ensure full employment. Under the classical gold standard, as the gold stocks of some countries diminished, those of others increased and, in theory, the effects on expenditure in the ones offset the opposite effects in the others. But now the authorities took monetary and fiscal measures according to their own judgement and, in countries that did not need to improve their trade balances, they were free to let their residents' spending increase with the reserves or to restrain it. They might even believe that a spending increase enough to balance trade would be inflationary, although it would in theory have been automatic with the gold standard. Keynes had argued that surplus countries should be obliged to increase their expenditure, but, at the time, that involved commitments the American negotiators could not make¹⁶⁴ and, later, the central banks of countries with persistent surpluses, notably Germany, would acknowledge no such obligation.

Yet during the time the Bretton Woods system functioned Europe had no depression, just mild recessions. Europe's governments did not put financial objectives and reparations ahead of economic recovery, as they had done after the First World War, and they began social programs to benefit the mass of their populations. As a consequence, their expenditures in the first years after the war exceeded the production capabilities of their economies and trade surpluses were confined to the US, which provided the necessary financing. They could, nevertheless, have subsided into depression and inflation had the US not provided financing and aid in kind to offset the low saving rates and shortages of essentials. Another possible cause of depression was the competitiveness of American industrial, agricultural and mineral exports, which could have prevented the initial recovery at any exchange rates. This was prevented by administrative controls over imports and high tariffs, which allowed the European economies to produce goods that were not competitive with those available from the US and at the same time to offset the lack of trade among themselves. Tariffs were high in the US, too; the weighted average of dutiable imports was 48 per cent until a round of tariff negotiations in 1947 led to them being lowered to a weighted average of 25 per cent in 1948.¹⁶⁵ European tariffs were lowered in step, but remained high enough for production and investment to increase. A third possible cause of depression, a special concern of the British government,

¹⁶⁴ Gardner, *Sterling Dollar Diplomacy*, 92–93; Solomon, *The International Monetary System, 1945–1981*, 32.

¹⁶⁵ Wilcox and United Nations Conference on Trade and Employment, *A Charter for World Trade*, 64.

was a depression in the US. Nothing worse than a short recession occurred there, but even a real depression would probably not have had much effect on Europe's economies, which exported little to the US at that stage. In time, as they increased their exports to the US, these economies might have been more affected, but, again, there was no depression and these economies were by then trading more among themselves and the rest of the world.

Yet, two years after the war it appeared that these possible causes of depression and the poverty might prevent Europe's economic recovery, which was further jeopardised by an exceptionally severe winter. It seemed to American and European officials that Europe's needs had been underestimated and the means by which the US had hoped to finance the recovery, namely the loans provided directly and through the World Bank, along with aid in kind and the indirect relief from not insisting, as it had done after the First World War, on repayment of loans, would not suffice. Even if the US officials had thought this financing would be insufficient, they could not have obtained more from their Congress then. But the efforts of the World Bank were slow and small compared to the urgency and scale of the needs arising from the destruction of infrastructure, housing and industrial capacity. Such industrial capacity as could be used or repaired had to be converted from war production, which, itself, needed investment and time. Shortages of essentials were common and often localised. There were also several million destitute refugees. The most urgent needs had to be met through supplies provided by the US and other countries, partly through the United Nations, and distributed directly to the recipients. Various essentials had also to be rationed until enough of the transport systems and production capacity had been repaired for markets to function. Since dollars and gold were scarce and there was no functioning institution through which international payments could be made, there could be no quick removal of administrative control over international transactions of all sorts and, hence, no multilateral trade. Bilateral exchanges between the countries of Europe did reduce some scarcities, but were minimal.

At this point the US increased its aid with the Marshall Plan (the European Recovery Program), which operated from 1948 to 1952. By then the Cold War caused enough alarm in the US for Congress to agree to the Plan. Communism was feared as an imminent threat; Russia occupied all of East and much of Central Europe up to parts of Germany and Austria, apart from which Albania and Yugoslavia¹⁶⁶ were communist. In West Europe communist parties were stronger than they had been before the War, both because they seemed to offer alternatives to the failures of the pre-war years

¹⁶⁶ Yugoslavia comprised what are now Bosnia-Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia and Slovenia.

and because their standing was high for having fought against fascism, especially in France, Greece and Italy. Roughly the same amount of \$13 billion was disbursed under the Marshall Plan as previously under all the other heads and, in addition to providing financing for investment, it provided the liquidity needed to create a limited multilateral payments system, the European Payments Union (EPU), set up in June 1950 with monthly clearance of accounts and with one dollar as the unit of account.

In the countries of Europe expenditure was kept up by the state, even after the worst shortages had been relieved and markets had begun to function where they did not before, and also by consumers as wages rose with income. After the experience of the years between the two wars and the Second War, itself, there was a common feeling that the causes of the social divisions of the past could and should be removed or mitigated. In most of Europe the state had begun or expanded programs of education and health care, as well as clearing slums, subsidising public transport to benefit the mass of the population and providing more unemployment benefits than before. Governments also tried to ensure full employment; in the UK, for instance, though not in Germany, it was made an explicit commitment. Wages could not, therefore, be lowered to lower costs and reduce consumption. Besides, in some countries, notably France, Greece and Italy, the membership of the communist parties was large and social discontent could increase it. Since protection against imports remained high, even after the first round of trade negotiations had resulted in some reductions in 1948, investment was not deterred by lack of demand or by foreign competition, but constrained by saving.

Hence, up to the mid-1950s there was, in addition to shortages of essentials, effective demand in monetary terms that could not be met entirely from the production of the countries, and these were expressed as demand for dollars to pay for imports, the "dollar shortage" as it was known then. It meant that the US trade balance with Europe was always in surplus and that, after the Marshall Plan, the dollars had to come from the US capital account or from military expenditure in Europe. They did not come from American financial institutions, which judged direct lending to European governments or firms to be mostly too risky, and the lending of the World Bank was on too small a scale. But American firms and individuals invested in buying existing or setting up new production capacity and in acquiring property. Then came military expenditure, notably on the Korean War from 1950 to 1953. Together they resulted in overall balance of payments deficits in the US and an accumulation of dollars in Europe.

Economic Growth and Liberalisation in the 1950s

Had current accounts been convertible these countries would have had bigger trade deficits and would have lost what little reserves they had, unless prohibitive tariffs were imposed on imports. This was shown by the attempt by the UK, in July 1947, to open the current account in keeping with its undertaking with the US, a step that caused an immediate loss of the UK's dollar reserves and had to be reversed quickly. In theory the authorities of each country could have reduced expenditure to the point that the trade and current account deficits could be financed from the aid received, which would have left the saving rate and tax revenue too low for the investment and social spending the governments considered were needed, which is to say that Europe was too poor to recover quickly enough or at all without external capital.

But all agreed on the desirability of multilateral payments and freedom of transactions on the current account and they brought them about step by step. By the end of 1958 most of Europe had current account convertibility, though restricted to residents of countries with the same, except for Germany and the UK, which had no restrictions, and by early 1961 most of these countries had fully convertible current accounts in conformity with Article VIII of the IMF, the others following in the next few years. None of this would have been possible had it not been for the US capital account transactions and the resulting deficits on the balance of payments. Because of them, Europe, as a whole, could run trade deficits with the US for the first ten years after the war and still accumulate reserves, although the US had overall trade and current accounts surpluses almost uninterruptedly until 1971.

This gradual liberalisation of international payments, if slower than originally envisaged, was in the spirit of the Bretton Woods agreements but was not matched by liberalisation of trade. From 1948 to the Kennedy Round of negotiations beginning in 1962, tariffs remained high and quantitative restrictions on the exports of low wage countries, notably the Cotton Textile Arrangements, became more restrictive. An International Trade Organisation (ITO) for the reduction and eventual elimination of barriers to trade had originally been included in the Bretton Woods agreements, along with the IMF and the World Bank, but it failed to get the approval of the US Congress and would almost certainly have been rejected by the British Parliament and perhaps by some other countries' legislatures.¹⁶⁷ An agreement on the principles of the ITO had to satisfy the imperial preferences of several European countries and the demands of low wage countries to be allowed to protect their industries and agriculture, whereas the US insisted on non-discrimination, rather than low tariffs, and some exceptions for its agriculture.

¹⁶⁷ Gardner, *Sterling Dollar Diplomacy*.

What the negotiators of the countries involved in 1947 could finally work out had too many compromises and complications for the US Congress or the UK Parliament to accept. But agreement had been reached on tariff reductions for a number of goods, mainly of interest to the high wage countries, though the level of protection still remained high. The weighted average of the US import duties, for example, was 25 per cent¹⁶⁸, apart from which there were specific restrictions on agricultural products. This agreement, the General Agreement on Tariffs and Trade (GATT), though not formally an institution to replace the ITO, became in practice the institution for trade negotiations and the settlement of trade disputes until the creation of the World Trade Organisation in 1995.

The 1950s were an early stage in the process of liberalising international economic transactions, but in Europe they were a time of fast growth with wages rising in step, low unemployment, the welfare state, few serious budget problems and moderate inflation. Germany had its *Wirtschaftswunder* and opened the capital account to residents in 1957, but France and Italy, among others, also grew fast. Japan's economy and wage rates grew even faster. Aside from Canada's currency, left to float from 1950 to 1962, changes of exchange rates and need for the IMF's help were rare, though the IMF provided loans to several low wage countries, including one to Japan in 1953. Yet, as fast as Europe's trade was, it was slower than the growth of its reserves. Almost all the increase of reserves outside the US from 1951 to 1964, from \$23 billion to \$52 billion, went to Europe, where the ratio of reserves to imports rose from 32 per cent to 41 per cent.¹⁶⁹ Hence, after 1949, when the UK and several other countries devalued, only one European country, France, devalued before 1961 and beside it only one other country, the UK, borrowed from the IMF, and in each case part of the cause was political. In 1956 both countries had been running trade deficits and also caused the Suez crisis, and in 1957-58 France went through an internal political crisis. Both obtained IMF stand-by loans in 1956 and France obtained another in 1958. France devalued in 1958 and again in 1959, though the UK kept its exchange rate.

Exchange Rates and Speculation

Europe's and Japan's rapid economic growth continued through the 1960s, but the lack of the coordination needed for the mutual consistency of the expenditure of the various countries eventually led to the system's fixed exchange rates being replaced by increasing reliance on changes of exchange

¹⁶⁸ Wilcox and United Nations Conference on Trade and Employment, *A Charter for World Trade*, 64.

¹⁶⁹ International Monetary Fund., *International Reserves: Needs and Availability*. Calculated from Tables 2&3, pp.383,386.

rates. During the 1950s some countries had repeatedly had to improve their trade balances by curtailing expenditure and reducing production below the limits of their production possibilities, but ran into trade deficits when they allowed production to return to the limits. Such a country could be declared to be in fundamental disequilibrium and be allowed or obliged to devalue. Both France and the UK were such countries, but devaluation was regarded at the time as a symptom of failure and their authorities opposed it. Others, Germany in particular, consistently ran trade surpluses. Without coordination of expenditure, a country of the first kind had two alternatives. One was to remedy trade deficits by keeping expenditure on goods below income from production, though that might reduce its own income from production and worsen the trade balances of other less competitive economies without necessarily affecting those of the more competitive ones. The other was to try to improve its competitiveness by devaluation.

But, once devaluation became a possibility, there was bound to be speculation in the form of converting holdings of that currency into currencies that were expected to be stable, and that could result in devaluation. If it did, those holding that currency avoided a loss if they converted in time. If the regulations permitted it, speculators could convert deposits in the country's financial institutions into foreign currencies or gold, both of which came, directly or indirectly, from that country's central bank, for financial institutions that made the conversion would usually pass the currency on rather than hold it until devaluation. Either type of speculation could deplete the reserves a country ordinarily needed to finance trade and to cover current account deficits during the time the authorities needed to restore surpluses by cutting back expenditure and, unless the country obtained enough financing for the speculators to lose interest or devalued by enough to make it seem that its balance of payments would be safe, a crisis had to follow. Since there was no chance of the currency appreciating, speculators ran no risk of an exchange rate loss if there was no devaluation. Still, the devaluation had to occur soon for the speculation to be worthwhile, for the speculator incurred costs; while waiting for the devaluation the money had to be kept as idle deposits or invested with low returns or with some risk. Moreover, if the authorities of the country concerned clearly wanted to avoid devaluation and could delay it, their chances of avoiding it would be greater.

Much of the need for IMF loans in the two episodes just mentioned arose from speculation. Devaluation was avoided in 1956 because the IMF loans were big enough, and France's two devaluations later on removed any risk of speculation on more devaluation for several years. These episodes also showed how inadequate the ordinary liquidity needed for trade was for preventing speculative transfers and how strict capital account controls

would have had to be to; both France and the UK had capital controls at the time, but their regulations were either too lenient or too easy to evade.

Hence, a country's reserves had to meet several needs. For present purposes they can be classified in roughly the way Clower and Lipsey classified them in 1968.¹⁷⁰ Two needs that could, in theory, be estimated, were that there always be the means to be sure that current account payments were made in time, i.e. liquidity, and that the reserves sufficed to give the time needed to improve the trade balance. A third need, which can for present purposes be subsumed under liquidity, was to accommodate non-speculative short term capital movements caused by international differences of interest rates. Finally, there was the need to cope with speculation, which Clower and Lipsey believed was 'not amenable to rational assessment'.¹⁷¹ They did not mention long term capital movements, such as direct investment abroad, perhaps because they considered them less volatile. They also stated that governments did not have clear ideas about how far they would go to restrain temporary non-speculative capital movements, since that entailed judging the costs and benefits of measures like changing interest rates; government officials thought in terms of 'shibboleths and conventions'.¹⁷²

Clower's and Lipsey's scepticism notwithstanding, Government officials had already devised a procedure for preventing speculation from leading to balance of payments crises and exchange rate changes; the IMF provided the financing needed to prevent such outcomes, on occasion in cooperation with countries whose currencies were not expected to depreciate. A country might not have the reserves to cope with speculation unaided, but it also did not have to. As long as some countries had accumulated redundant reserves, the means to counteract speculative movements of funds were there. What government officials had devised was a way of using these redundant reserves in cooperation with the IMF to compensate for the shortages of other countries as the need arose. Prudence might still require that the latter have some reserves, but, otherwise, the reserves such a country needed for the times when its currency was not expected to be devalued could be less, enough for the first three needs. Sohmén gave the analogy, 'What is needed to provide adequate protection against fires is not so much an average-sized pot of water in every house, but sufficiently wide pipes that pump the required liquidity (whose total need for this purpose will then be quite moderate) to wherever it happens to be needed at any moment, plus a reasonably efficient fire brigade.'¹⁷³

¹⁷⁰ Clower and Lipsey, "The Present State of International Liquidity Theory," 587–88.

¹⁷¹ Clower and Lipsey, 588.

¹⁷² Clower and Lipsey, 587.

¹⁷³ International Monetary Fund., *International Reserves: Needs and Availability.*, 30.

Countries with redundant reserves might consider cooperation to be in their interest and could let their central banks buy the currencies that speculators were selling in anticipation of devaluation if the IMF's means did not suffice. Apart from wanting to keep the system functioning, their authorities might believe that a competitor would become more competitive than desirable by devaluing or that their own countries might have to revalue, which would lower profit margins and, perhaps, reduce competitiveness. They tried at times to deter speculative deposits by non-residents using expedients like requiring banks to hold reserves of 100 per cent against them, preventing interest payments, charging fees and requiring that the deposits be held for fixed periods without interest. An alternative that central banks had to buying the currency was the swap, whereby they placed for each other's use equivalent amounts of their own currencies, the same amounts having to be returned at a specified time. This eliminated the losses central banks and their countries would have incurred were another country's currency to be devalued after all.

In principle, a group of countries could cooperate in this way without the IMF, but, in practice, the IMF was almost always involved. High wage countries were not necessarily willing to lend to a country with balance of payments problems without some assurance that the authorities of that country were doing what they could to improve the trade balance and to acquire the means of repaying the loans, and providing this assurance by working out the program of expenditure restrictions and other measures was what the IMF was created for.

It was simpler, therefore, to increase the financing the IMF could obtain and leave to it the organising of cooperation of a group of countries to provide more means when necessary. IMF quotas, which were reviewed every five years, were increased in 1959 and several times later and in 1962 ten high wage countries agreed on a General Arrangement to Borrow (GAB), which enabled the IMF to borrow from them at short notice, and nine countries made arrangements for swaps. The efficacy of such combinations of the IMF, with the additions to its financing capacity, and financing by other countries as needed was demonstrated in 1964, when the UK's government decided not to devalue although it seemed it would have to. At first, in April, the IMF provided \$1 billion using the GAB and then a group of ten countries added \$3 billion in December. In 1965, after its second quota increase, the IMF provided \$1 billion more. It demonstrated that, if countries were agreed that a speculative movement of money should not oblige a country to devalue, the mechanisms they needed had been put in place.

Dismantling the Bretton Woods System: Exchange Rates

Had governments and their advisers had been more willing to coordinate trade balances the Bretton Woods system could have gone on indefinitely. But the desire to coordinate was not there, although the need for it grew as volatility increased. Rather, the doctrine that exchange rates should be flexible and, ideally, should float, i.e. be determined solely by the market, was becoming popular among economists and seemed to show that markets, left to themselves, would obviate any need for action by governments.

Volatility rose mainly because of the increase of dollar holdings outside the US. Almost all the time the Bretton Woods system was in place the liabilities of the US increased, mostly because of deficits on the balance of payments (official settlements) and despite almost uninterrupted surpluses on current account. As Europe prospered and liberalised and as its dollar holdings grew, its private financial institutions and firms held dollar balances more readily and European currencies became increasingly convertible and hard. The growing amounts that financial institutions in Europe held of dollars and of European currencies outside their countries gave rise to an international financial market, the Eurodollar market, a source of financing for speculation, as well as for commerce and investment. Reserves that would have given a country time to improve its trade balance if there were no speculation no longer sufficed. Reserves did not accumulate evenly either; the countries that most often had trade deficits were likely to accumulate the least and those with perpetual surpluses the most. Foreign investment could help the former add to their reserves, but there was no assurance that the net amount would be enough, or even positive. Cooperation against speculation had, therefore, to be prompt and more frequent.

By itself this need not have caused a problem. It had already been shown, for example by the support for the UK in 1964, that the mechanism was there. So were the means, for the funds available to the IMF and the countries with redundant reserves for countering speculation, including holdings of other hard European currencies, had increased too. Their prompt, routine use would probably have made such speculation negligible. Speculation could have become stabilising as it had been with the gold standard before 1914, when speculators were confident that the authorities would act to preserve their gold prices.

But cooperation was hindered by political complications, which need not be considered here but which arose in great measure because it was mostly the same countries that had repeated deficits or persistent surpluses. They would not have occurred had trade balances been random, countries being as likely to have surpluses as deficits at any time, for countries would have been alike in this respect. Since they did occur and a permanent

arrangement by which countries with redundant reserves helped the others, however simple technically, was never envisaged, the sole means for coping with persistent imbalances seemed to be expenditure restraint accompanied if need be by exchange rate changes. As speculation became more frequent it seemed that countries should be allowed to change their exchange rates more easily and more often. The system of constant exchange rates was modified to allow discrete changes more often, the so-called pegged rates, and to allow greater variation around par and central rates. Yet speculative capital flows continued and made it seem that a set of exchange rates had to be found that would be stable for some time.

Despite the repeated changes, no set of stable exchange rates was found. In 1967 the UK devalued after running bigger trade deficits than before and in 1969 France devalued as well. Immediately after that Germany, with its perpetual trade surpluses, let its currency float for a while and then pegged it 9.3 per cent higher relative to the dollar. Countries with deficits also restricted their expenditure to avoid having to devalue, though it was only a stopgap that did not substitute for cooperation by the surplus countries in increasing their expenditure. Exchange rates were changed several times more before Europe finally abandoned fixed exchange rates in 1973, yet mostly it was the same countries that devalued and the same countries that revalued, from which it could have been surmised that changing exchange rates was not even a stopgap and that any improvement of trade balances after devaluation was the result of the expenditure restrictions that always accompanied it. The self-perpetuating process of more frequent exchange rate changes and more money devoted to speculation continued.

A concerted attempt to restore stability was made in December 1971 with the Smithsonian agreement, by which the nearly all exchange rates of the high wage countries were changed, the dollar being devalued relative to the others, excepting the French franc and the British pound. Only by such concert could the dollar be devalued against other currencies when their exchange rates were set in terms of the dollar. It had to be matched by an increase of the price of gold, which France and the US had agreed the previous month would be raised to \$ 38 an ounce. Another change that it was hoped would add to stability by making it easier for countries to avoid changing parities was a widening of the bands around which the actual exchange rates could fluctuate from 1 per cent to 2.25 per cent around their parities or central rates, though the EEC countries, which wished to integrate their economies more, chose to limit the fluctuations between their currencies to 2.25 per cent – the “snake” within the Smithsonian “tunnel”. The new exchange rates and wider bands altered little; the UK devalued six months later, the US devalued against gold to raise the price to \$ 42.22 an

ounce fourteen months later and some countries left and sometimes re-joined the snake in the following years.

At the time the theoretical arguments for floating exchange rates were becoming the orthodoxy in the US, and its authorities were, accordingly, growing reluctant to take measures to reduce additions to dollar holdings abroad if, as theory seemed to imply, letting the exchange rate float would bring about equilibrium. For example, fifteen months later, at a meeting of fourteen countries in Paris in March 1973, they declined a number of proposals for such measures. Despite the aftermath of the Smithsonian agreement and despite the trade balance of the US, which is what the exchange rate would in theory have affected, having been in surplus along with the current account until the agreement, American officials continued to act as though the US balance of payments deficits were caused by the undervaluation of other currencies.¹⁷⁴ The US had never had a deficit on the trade or current account before then, and it has had deficits almost uninterruptedly since. It seems, too, that the European countries' officials involved in the Smithsonian agreement, who showed no sign of having anticipated the American trade deficits to come, were willing to accept greater trade deficits for themselves as a whole.

Dismantling the Bretton Woods System: Gold and the Special Drawing Rights

As exchange rates became ever more flexible gold's functions as a reserve and standard were eliminated. At \$ 35 an ounce there had been no prospect of increasing the total supply of gold fast enough to allow the US money supply to grow as it had been doing and to keep up with the world's demand for gold, especially if other countries' central banks and the general public chose to convert much of their dollar holdings, and there was still no such prospect at \$ 42.22.

Doubts that the gold exchange standard could continue without change were not new; they had already been expressed in the late 1950s. For example, Triffin had begun to argue that the balance of payments deficits provided reserves or liquidity necessary for the growth of world trade¹⁷⁵ and warned of the dilemma, '... if the US corrected its present balance-of-payments deficits, the growth of world reserves could not be fed adequately with gold production at \$35 an ounce, but that if the US continued to run deficits, its foreign liabilities would inevitably come to exceed by far its ability to convert dollars into gold upon demand and would bring about a "gold and

¹⁷⁴ Solomon, *The International Monetary System, 1945-1981*, 239.

¹⁷⁵ Triffin, *Tomorrow's Convertibility: Aims and Means of International Monetary Policy*.

dollar crisis".¹⁷⁶ He believed that this would cause problems that could end in a worldwide financial crisis, a belief that was widely held at the time, and proposed creating an international reserve asset to supplement the dollar. Already in October 1960 the price of gold on the London market rose briefly to \$ 38 – 40 per ounce. This was just before the US presidential elections and John Kennedy, who was campaigning to be elected, felt obliged to declare that he would defend the official price.

As long as they had not agreed on how the Bretton Woods system should be changed the high wage countries cooperated in trying to keep gold as a reserve asset. In particular, keeping the gold exchange standard at \$ 35 an ounce for gold required that some of the deficits of the US balance of payments had to be reduced. A first step was taken already in 1960 with the reduction of some types of US military expenditure abroad. It was rescinded the next February but was followed by a succession of measures to deter or control capital transfers abroad by US residents and even surcharges on imports. Central banks avoided buying so much gold from the US as to reduce the reserves to the minimum compatible with the money supply, though they also wanted to have their own shares of the gold stocks. An expedient, reminiscent of the years before 1914, by which they could do both was the gold pool for buying and selling gold at the official price, to which eight countries contributed. It began in 1961 and became a formal arrangement in 1964.

Being a buffer stock scheme for gold and dollars, it was bound to fail eventually, when the supply or demand for gold or dollars exceeded what it could accommodate. This happened in March 1968 after devaluation of the British pound in November 1967 set off an increase of gold buying that reduced by an eighth the reserves of the contributing countries, which France had left in 1967.¹⁷⁷ Around then the US Congress removed the gold cover requirements for Federal Reserve notes, an alternative President Kennedy had considered in 1961 to be able to use the gold stock for keeping its price constant and had rejected for political reasons.

Removing the gold cover requirement ended the gold exchange standard. Instead of using the gold reserves to keep the market price at \$ 35 an ounce, i.e. as a buffer stock, the monetary authorities of the major economies agreed to separate their reserves of gold from the market, keeping them among themselves as reserve assets at the official price and not profiting from arbitrage when the market price was higher. On the one hand, they wanted, to keep the price of gold constant as though the standard were still in place; they did not want it to rise, because that would have been devaluation of the dollar, nor did they want it to fall, because it was a reserve

¹⁷⁶ Triffin, *Gold and the Dollar Crisis: The Future of Convertibility*, 62.

¹⁷⁷ Solomon, *The International Monetary System, 1945-1981*, 119.

asset for all countries. On the other hand, they wanted to preserve their gold reserves as much as possible and judged using them to keep the price constant would sooner or later result in losing a part to the market. Thus were eliminated the two main prerequisites of a gold standard, the obligation of the monetary authority to limit its currency issue according to a relation to the stock of monetary gold set by law and the obligation to buy gold from or sell it to the public at prices close to the official price. In its place was a two tier gold market, the old official price for the transactions of the monetary authorities and a market price.

By 1968 the Bretton Woods system had been dismantled and a new international reserve asset had to be found or devised. That the supply of gold could not be adjusted as the relevant authorities judged desirable might have been thought an advantage in the nineteenth century, but was now considered a drawback. Some substitute for or supplement to gold, apart from the dollar, was felt to be necessary. A process of discussions, studies and negotiations had already begun in the early 1960s and culminated with agreement on the creation of the Special Drawing Right (SDR) by the IMF in 1968 as a supplement to gold with its value set in terms of gold. The Articles of Agreement were amended accordingly with effect from July 1969 and the Board of Governors of the IMF decided that September to create SDRs equivalent to \$9.5 billion over three years beginning in January 1970.

Now the SDR had come into existence and the dollar was no longer convertible into gold there had to be a procedure for deciding on what quantity of SDRs to create, especially if, as Triffin and others believed, the demand for the dollar as an international reserve asset had depended on its convertibility into gold. In July 1970 the IMF held a seminar in which a number of prominent economists from different countries took part because it would "... benefit from close contact with the academic community ..."¹⁷⁸ and published the proceedings as a book, "International Reserves: Needs and Availability". No firm conclusions or estimates resulted, nor were any asked for. But there was agreement that the world had a determinate need for reserves and the means of assessing it ought to be found. Polak, who presided over the seminar and whose macroeconomic model has been used routinely by the IMF in formulating the economic programs of countries asking for its financing, stated that the supply of international reserves "... had been the result of random, unpredictable, and uncontrolled forces – the output of gold mines, the hoarding of gold, wheat production in the Soviet Union, the balance payments of the United States. ... No specific arrangements have been made to determine the supply of reserve currencies but there is general agreement that this supply cannot in the future continue

¹⁷⁸ International Monetary Fund., *International Reserves: Needs and Availability.*, ix.

to rise in the manner in which it did in the fifties and the first half of the sixties.¹⁷⁹ A Staff Paper written in 1966 and included in the proceedings had, after describing the many problems in estimating the world's reserve needs, made a tentative estimate from past trends that they should increase by \$ 3 billion a year.¹⁸⁰

Many points were at issue, among them, how were the amounts to be determined and by whom, who was to emit the reserves, were they to be transferable between the monetary authorities of countries or were they to be credits drawn on and repaid to the emitting institution, if transferable would the public be allowed to hold them and what would the functions of the dollar and gold be? Neither the possible answers nor the discussions and negotiations need be recounted here, for in the end gold was demonetised, the dollar became the main currency for international payments and reserves and exchange rates were left to float. The SDR, which was meant to be the new form of international reserves, has never been of much account, except for accounting technicalities of the IMF. Its only other issue has been of SDR 161 billion to countries of the former Soviet bloc in 2009 to help mitigate the effects of the financial crisis in the US and Europe, which brought the total to just over SDR 204 billion, equivalent to about \$ 309 billion then. According to the IMF, the world's total reserves, excluding gold, were equivalent to roughly SDR 5,449 billion, at the end of 2009 and SDR 8,269 billion, roughly \$ 11,800 billion, at the end of 2017. Since both the US and the main European countries acting together had enough votes in the IMF to prevent issue of SDRs an adequate increase of the number of SDRs required agreement, which did not come about.

With the reserve function of gold limited by its official price and the and the issue of SDRs kept so small, the greater part of almost any country's reserves consisted of dollars and revaluing its currency relative to the dollar meant a loss in terms of that currency. At the same time US officials maintained that the US balance of payments deficits were caused by others, especially the European countries, and accordingly proposed that countries whose reserves were excessive according to some agreed indicator should take measures to reduce their surpluses, it being understood that the measures were revaluation, not expenditure increases. Apart from the addition of an indicator, this resembled Keynes's proposal for the Bretton Woods system. But the circumstances were different. Now countries holding excessive reserves were being asked to revalue against the dollar when dollars constituted the greater part of their reserves. That they should hold their dollar reserves was part of the proposal. Other countries, too,

¹⁷⁹ Polak, *Money-National and International.*, 510.

¹⁸⁰ International Monetary Fund., *International Reserves: Needs and Availability.*, 412.

could expect their dollar holdings to depreciate relative to other currencies. To protect the value of their reserves they had to be able to convert dollars into gold or SDRs, which were at the time denominated in terms of gold, neither of which did the US have enough of.

If the US had had enough gold reserves and the IMF had issued enough SDRs, the monetary authorities of the various countries could have held the three assets in the proportions they liked. As it was, the US could not allow convertibility. Instead, its monetary authorities had made their objective the demonetisation of gold, that is, the end of its use as an international reserve asset and as a monetary standard. Triffin remarked in 1978, 'The brunt of U.S. official policies, however, remained to phase out gold, but not dollars, from international reserve creation. The American negotiators of the SDR agreement desperately tried to shape it in such a way as to make it "better than gold, but not as good as the dollar" ...'.¹⁸¹

Gold could have continued to be used for settlements between central banks at market prices, which is what the French authorities proposed at one stage, but it seemed at the time that the SDR would replace it. In August 1975 the Group of Ten countries agreed that for two years their total stock of gold, including the gold at the IMF would not be increased and that they would not fix the price of gold. The IMF would sell one sixth of its gold on the market to provide concessional loans to the developing countries and another sixth would revert to the countries at the original prices. At the same time the Interim Committee, which had been set up by the IMF in 1974 as part of the process for making the main decisions on monetary reform, agreed that the IMF's Article IV be amended to let countries leave their currencies to float, with the possibility for groups of countries to cooperate on their mutual exchange rates, and to forbid denomination of currencies in terms of gold, though denomination was allowed in terms of the SDR, the value of which was still fixed in terms of gold. Gold had been demonetised and replaced by the US dollar, not the SDR, and it had become apparent that the acceptability of the US dollar did not, as Triffin and Kenen feared, depend on its convertibility into gold.

Dollar Reserves, US Passivity and the Evolution of International Finance

Before countries began accumulating dollars because of US deficits on the current and trade accounts a common answer to the question as to how the need for reserves could or should be determined was that the US provided reserves according to other countries' needs. In other words, it was mostly passive. This was especially the answer of those who advocated making the dollar the main or only international reserve asset and means for

¹⁸¹ Triffin, *Gold and the Dollar Crisis: Yesterday and Tomorrow.*, 8.

making payments, for they had to argue that the US had not used its privilege of being the issuer of dollars unconstrained by the gold standard for abnormal economic gain. Even if it had been passive, though, there was no assurance that it would remain so. So, another argument put forward by McKinnon concluded that there could be no such gain.¹⁸²

In the following it is shown that these arguments do not take proper account of the components of the balance of payments. It is easier now, knowing how long the US trade and current accounts have been in deficit and the magnitude of the deficits, to spot the components of the balance of payments that were not passive but determined by the choices of the US state and private sectors than it was then, when the US trade and current accounts were in surplus. Nevertheless, there were complaints in Europe even then, especially in France, that the privilege was 'exorbitant' and new arguments, equally lacking in cogency, have been devised more recently to deny the 'exorbitant privilege'.

Passivity in the sense used here presupposes that the rest of the world's demand for dollar reserves at any moment is given. If it is determinate, therefore, an econometric estimate must specify it as a function of which the arguments are chosen according to the assumptions as to how the demand is determined. Then, with some simplification, the form can be determined econometrically from the actual reserves. Several models were devised with different assumptions to determine this function. In an article of 1960 Kenen assumed that countries held gold and dollars in proportions determined by the interest on the latter and the risk of a rise in the dollar price of the former.¹⁸³ Another type of model was meant to explain how the need for reserves grew as trade increased, the US merely supplying the dollars in response. Several were devised with different assumptions about the relations between the reserves countries desired and the various quantities connected to these countries' foreign trade.¹⁸⁴ In these a random element was assumed to allow statistical estimates of these relations. At the simplest, the level of desired reserves was assumed to increase steadily with time and actual reserves to be increased or decreased in proportion to how much they differed from that, with the random element adding to or subtracting from that. Other models related desired reserve levels to exports, imports, the trade balance and the variability of these quantities, sometimes trying to take the opportunity costs of holding reserves into account.

¹⁸² McKinnon, *Private and Official International Money: The Case for the Dollar*.

¹⁸³ Kenen, "International Liquidity and the Balance of Payments of a Reserve-Currency Country."

¹⁸⁴ A survey is given in Williamson, "Surveys in Applied Economics: International Liquidity," sec. II.

At first sight the results of the statistical estimates made the models plausible, for nearly all the statistical relations were positive and significant, even those for imports, which in most of the models should have been negative. Williamson took this to be evidence for them,¹⁸⁵ though it is better to argue that the same positive significant relation for different quantities and different models is a sign that a relevant element has been left out, that the specifications of the estimates are wrong. For example, if reserves grew independently of the quantities in the estimates, the growth of trade would be expected to yield positive and significant coefficients. It would have to be shown, then, how the US was the passive source of the growth of reserves, otherwise the conclusion would have to be that the US was not passive and the actions of its residents also determined the creation of the world's dollar reserves.

If the US had only provided the reserves passively, the mechanism would have had to have been providing loans in dollars on demand, though little, if any, of that demand could have been of a kind that resulted in accumulation of dollar holdings and reserves. Loans to finance imports do not result in the accumulation of reserves. Nor do loans to buy equity in the US or tangible assets, like production capacity and property. US dollar loans to consumers in Europe and elsewhere that were not spent on imports from the US would not normally have been long term loans and, since they had to be repaid, would have added little to reserves. Their total would have been too small to need to be taken into account and were ignored in the arguments discussed here. Some loans for consumption were foreign aid in the form of balance of payments support or for emergencies and would all have been spent on imports. All that is left is loans for acquiring financial assets, including equity, and loans for creating productive capacity or acquiring tangible assets outside the US and they could have been to US residents or to foreigners. To the extent that they were for investing in creating productive capacity and led to imports from the US, they were financing of imports and did not add to reserves and to the extent they were loans to US residents, whether for acquiring financial or tangible assets, they were not passive.

Passive lending cannot explain why, in practice, some countries acquired reserves beyond what could be considered reasonable needs, one of the practices that made the Bretton Woods system untenable. Reserves yield little or nothing¹⁸⁶ and adding to them has a cost, which is the goods and assets the country could have bought instead. H. G. Johnson made the point in 1966, 'In real terms, a surplus in the trade balance used to finance an accumulation of reserves involves a transfer to the rest of the world; and,

¹⁸⁵ Williamson, 695.

¹⁸⁶ Fleming, "Reserve Creation and Real Reserves."

conversely, the use of reserves to finance the trade deficit involves receipt of a transfer from the rest of the world.¹⁸⁷ Johnson wanted to show that floating exchange rates had the advantages of reducing the transfers from some countries to others and, perhaps, of preventing a welfare loss by requiring fewer reserves than fixed exchange rates.

In theory, reserves are held for precautionary reasons; just as individuals may want to hold liquid assets, countries hold reserves as protection against fluctuations and unforeseen events and, if the country's authorities can choose how much to hold, holding more than necessary is irrational. Nevertheless, some countries consistently accumulated reserves through trade surpluses, were "reserve sinks", and others had to cope with repeated deficits. The British economist, Kahn, pointed out already in 1960 that some countries had excessive reserves and French officials, for instance the finance minister, Giscard d'Estaing, made the same point repeatedly.¹⁸⁸ Since the reserve sink countries would not increase their expenditure to allow those that had repeated trade deficits to obtain surpluses, creating a new reserve asset for the world as a whole would merely have allowed the deficit countries to increase expenditure for a while or to pay off some foreign debt, but not to increase their reserves for long. Neither the IMF nor any countries had means of inducing reserve sink countries to stop accumulating reserves. Under the gold standard, in theory, countries regulated themselves by letting the money supply increase when the gold reserves increased, the consequent increase of demand causing a worsening of the trade balance directly or through inflation. If international reserves had the same effect on the money supply without gold being the standard, the effect might have been the same. But there seems to have been no limit to how long the monetary authorities could go on sterilising reserves, which is to say, their ability and willingness to restrict expenditure was not affected by their balance of payments surpluses or reserves.

Models like those described above, in which the US supposedly supplied dollar reserves to meet demand, do not take the disparities between reserve sinks and the others into account. Apart from the high wage countries that had balance of payments problems, there were also low wage countries that were often short of reserves or only able to obtain more at excessive economic and social costs. Implicit in these models is the assumption that the monetary authorities of all countries apart from the US have the same criteria for desirable levels and can satisfy them. If there were such criteria there would have been no need for the models; inquiries with the relevant authorities should have sufficed to find out what they were, unless they were

¹⁸⁷ Johnson, "The Welfare Costs of Exchange-Rate Stabilization," 512.

¹⁸⁸ Giscard d'Estaing, "La Politique Monétaire Internationale de la France."

state secrets. Statistical exercises and *a priori* reasoning of the kind described above leave unexplained how the criteria were chosen, who in the countries chose them and why the institutions and political leaders abided by the choices. What those who use statistical methods to estimate coefficients of types of behaviour prescribed in models assume, in effect, is that decisions about reserves are not made with deliberation but are reached by some kind of rational, but subconscious process.

Had the authorities of the European countries with persistent surpluses, notably Germany, been asked why they made their countries reserve sinks they would have answered that they had no better choice. They believed that, with full employment, greater demand would have been inflationary. Before the Eurodollar market and, later, international financial markets had evolved European exporters converted their foreign exchange receipts into their domestic currencies and, as their banks deposited the foreign exchange with the central banks and withdrew what was demanded for external payments, the excess that remained was sterilised and deposited in the reserve account. Not much else could be done. Firms and individuals normally kept their bank deposits in their countries of residence and foreign investment by European firms was not common, usually undertaken by bigger ones for the extraction of minerals. It was in the 1970s and 1980s that international financial markets evolved and international payments were liberalised to the point that international financing became routine and firms became open to holding deposits in various currencies and several countries.

As firms and individuals, exporters in particular, became free to place their foreign exchange in their domestic or foreign banking systems as they preferred and as the practice of producing components of goods in different countries spread, such financial possibilities became necessary. One effect was that central banks and the state became obliged to buy what foreign exchange they needed for reserves and other uses from the financial markets, where firms and individuals had their deposits, but they also became able to place foreign exchange on financial markets to get better yields. Nowadays a country with a hard currency can buy and sell on the financial markets to have reserves composed of the amounts of other hard currencies and gold it chooses, even if it has persistent current account deficits. For example, the UK current account has been in continually in deficit but, as long as its reserves suffice for all but the most extreme events, the market does not worry. That is assurance enough for foreign investors, including individuals who acquire property in the UK.

This evolution of the Eurodollar market and the present world-wide financial system could not have been predicted from the argument that the US provided liquidity passively and its implicit neglect of the differences between

countries in their reserve holdings. For example, in an article that is often referred to, originally published in the weekly, *The Economist*, in 1965, Kindleberger, Despres and Salant, asserted that, ‘... the USA has provided the world with liquid dollar assets ... by capital outflow and aid exceeding its current account surplus ...’,¹⁸⁹ and the ‘deficit’ has largely reflected a ‘... process, in which the USA has been lending, mostly at long and intermediate term, and borrowing short.’. Because the preference for liquidity is greater in Europe and because the US has a superior capital market, the US has supplied financial intermediary services to the extent that ‘... its loans to foreigners are offset by foreigners putting their own money into liquid dollar assets ...’; ‘... it supplies liquidity to foreign asset-holders, who receive less for placing their short term deposits at home. ... just as differing comparative costs create the scope for mutually profitable trade in goods’.¹⁹⁰

A priori it is implausible that such intermediation could have amounted to much. It can be seen from the point at which the analogy of comparative costs is misleading; in trade the goods go to different markets, whereas the proceeds from borrowing long term are supposed to become short term deposits in the same financial market as the loans with a consequent loss on the interest. This would normally occur only if the deposits were to be used quickly, which is when the liquidity is useful. Long term loans would, therefore, have been financing for purposes other than liquidity and short term deposits would have been working capital and small. If, as Kindleberger et al. seem to imply, the short term deposits of foreigners in the US were of amounts comparable to the long term loans, the borrowers could not have been the same as the depositors and the long term borrowing was not to obtain liquidity.

Later they refer to ‘liquid saving in Europe ... matched by European borrowing in the USA ...’ and say savers and banks ‘... typically want liquidity in their own currencies ...’ and the central bank of their country must hold the dollars or gold.¹⁹¹ This seems to mean that savers and banks who already had liquid savings borrowed in the US and then converted some or most of the loan proceeds into liquidity in their own currencies. They were either adding to their liquidity, though it is not explained why they should do so when they already had it, or they were placing their liquid savings as short term deposits in the US, presumably as collateral for long term loans in dollars, in which case they would again have lost on the interest.

In claiming that the excess of the capital outflow over the current account surplus largely reflected long and intermediate term lending by the US,

¹⁸⁹ Kindleberger, *International Money: A Collection of Essays*, 43.

¹⁹⁰ Kindleberger, 43.

¹⁹¹ Kindleberger, 49.

Kindleberger et al. ignored the question as to which items of the balance of payments were passive and which active. Over the period 1958-65 the cumulative current surplus was \$ 20.3 billion and the capital account deficit, excluding reserves, was \$ 30.2 billion. In the capital account government spending and direct investment, which were ignored,¹⁹² must be considered active items and they came to \$ 9.0 billion and \$ 13.4 billion respectively. Private long term lending came to \$ 8.5 billion, and, as argued here, little of that could have been used for intermediation. To this can be added private short term lending to give a lending total of \$ 14.4 billion. How much of that went to foreigners cannot be known with certainty, but it is unlikely that foreigners could at that time have accounted for much. Kindleberger et al. did say that the excess of the capital outflows over the current account surplus `... supplied goods and services to the rest of the world ...',¹⁹³. According to them it was a minor item and, if it meant, as it seems, financing of exports, it was not part of the intermediation. They also asserted that the excess `... supplies loans and investment funds to foreign enterprises which have to pay more domestically to borrow long-term money and which cannot get the amounts they want at any price.'¹⁹⁴. This must also have been a minor item to the extent that it was distinct from the previous one, for it was not intermediation. It gives the false impression that the countries of Europe could not finance their growth because of their preference for liquidity, and this at a time of unexampled economic growth. No evidence was given for the assertion.

Hence US residents must have accounted for the greater part of the lending on capital account and that means that almost all the capital outflows consisted of active items in the sense of not being determined by foreign market demand but by decisions of US residents. Another active item, which Kindleberger et al. overlooked because they confined themselves to the aggregate current account, was military expenditure, an item of the current account that amounted to \$ 21.6 billion over 1958-65, as compared to a trade surplus of \$ 34.9 billion. Foreign commercial bank and other foreign short term deposits over this period amounted to \$ 6.6 billion and can only be explained as liquidity generated by the active expenditure of the US, not borrowed by the depositors.

The following years, 1966-70, provide more evidence against the assertions of Kindleberger et al.. Private long term lending, having been an outflow of \$ 8.5 billion in the years 1958-65, became an inflow on the capital account of \$ 9.1 billion because of the measures taken by the US authorities. This was not foreigners lending to the US, but US residents bringing their own funds back. Despite the change of long term lending, both foreign

¹⁹² Dudler, "Book Review."

¹⁹³ Kindleberger, *International Money: A Collection of Essays*, 43.

¹⁹⁴ Kindleberger, 43.

commercial bank and other private short term deposits in the US increased to a total of \$ 14.6 billion. Evidently the authorities, who had more immediate and detailed access to information on external financial transactions than academic economists, had seen that the long term loans had not been for financing foreign liquidity. Long term loans from the US could not, therefore, have been sources of the dollars abroad or deposited by foreigners in the US in that or the earlier period. At the same time the increase of foreign deposits in the US indicate that the dollar holdings of foreigners continued to increase. One reason for the increase was the decrease of the trade surplus, especially relative to the active items. It had been roughly equal to the sum of military expenditure and direct investment in 1958-65 and was now less than either. Government expenditure was also higher in this period. Thus, the active items continued to supply dollars, but the trade surplus absorbed less. Strange remarked in 1976, 'As dollars accumulated in the hands of surplus countries like Germany and Japan, the political asymmetries of a system that made it especially easy for American companies to acquire foreign subsidiaries and for US troops to live and fight abroad became more glaringly obvious.'¹⁹⁵

Seignorage and "Exorbitant Privilege"

It was at about the time when Kindleberger et al. made their assertions that de Gaulle, President of France, objected to the "privilege" the US had of being able to pay with its own currency as long as countries and markets went on accepting more dollars. It allowed seignorage, i.e. the ability to obtain goods and make investments at no cost to itself as a country and to service external debt with more debt, which the finance minister, Giscard d'Estaing, called in 1965 an "exorbitant privilege".

Later McKinnon, who also advocated the dollar as the international reserve and means of payment, tried to refute the contention that such seignorage could occur. He agreed with Kindleberger et al. that the reserves of Europe and elsewhere were created by borrowing from the US and asserted '... the outflow of long-term capital ... directly reflects the asset-preference functions of official and private foreigners. In order to build up their liquid-asset positions, foreigners borrow long in New York.'¹⁹⁶ "Liquid-asset positions" can be taken to be or to include the reserves of the models referred to above, for McKinnon adds that this '... process of financial intermediation more than offsets the current account surpluses to permit the international money supply to grow at or close to the desired rate.'¹⁹⁷ Here the word "desired" may mean equilibrium, which, as Dudler pointed out in his book review, is at odds with

¹⁹⁵ Strange, *International Economic Relations of the Western World*, 320.

¹⁹⁶ McKinnon, *Private and Official International Money: The Case for the Dollar*, 14.

¹⁹⁷ McKinnon, 14.

the European monetary authorities' deliberate sterilisation of dollar assets and restraint in converting them into gold.¹⁹⁸ If "desired" does not mean equilibrium, there being no explanation as to why or by whom that rate should be desired, the word is gratuitous and means no more than that the rate is the market outcome under those conditions.

Of the possibility that the US could '...create credit "costlessly" and to obtain real goods and services in exchange', McKinnon asserted it could not be.¹⁹⁹ He referred to this possibility as "unrequited seignorage" and denied that any necessarily accrued to the issuer of the international money or that there could be any if the financial system was competitive. His reasoning was that true seignorage occurs when the state '... appropriates real resources through the issue of noninterest-bearing (*sic*) debt, through the suppression of interest rates, or through the exercise of reserve requirements.'²⁰⁰ Assuming that the central bank is independent, as the Federal Reserve of the US always has been, and does not increase bank reserve requirements and that the financial system is competitive, the state can only borrow in competition with private borrowers, and, if it pays less interest because of lower risk, that is not seignorage.

McKinnon's argument does not remove the possibility that the US created credit at no cost to itself, as a whole, and obtained real goods and services in exchange, for it overlooks the question of to whom the interest is paid. To put it differently, if the definition of seignorage is not confined to the state, *qua* government, but is extended to the country, the US obtained seignorage through foreign investments and military and government spending abroad. Neither the military nor the government spending necessarily incurred any interest costs, but, even if they did, when the state borrows from the central bank and uses the money to pay for goods from other countries, the interest the state pays accrues to its central bank whilst the foreign sellers receive that country's currency as notes or bank deposits, which do not necessarily bear interest. Similarly, a firm or individual borrowing from a US bank to buy shares in a firm, set up new production capacity or acquire property in another country pays interest to the US bank, whilst the payments abroad of the firm or individual become bank deposits or currency notes held by banks or central bank in the other country. In either case, if the bank deposits of the recipients of the dollars yield interest, perhaps as Eurodollar deposits, it is not the US that pays it. In the same spirit Kindleberger argued that the countries of the British Commonwealth lost nothing by being obliged to keep their reserves in London in sterling since

¹⁹⁸ Dudler, "Book Review."

¹⁹⁹ McKinnon, *Private and Official International Money: The Case for the Dollar*, 4–5.

²⁰⁰ McKinnon, 20.

they could borrow against their reserves.²⁰¹ He would have been right if the reserves yielded interest comparable to the borrowing costs. Otherwise, it was more business for the banks of London.

Seignorage occurs when residents of one country use its currency to pay for the acquisition of goods, services or assets, tangible or not, from non-residents outside the country at no cost to the country as a whole. Thus, if the country reduces its payments for other purposes by the amounts paid for the acquisitions there is no seignorage. There is also none if these payments were made without a corresponding reduction of payments for other purposes, but the currency used for that is used by non-residents to acquire income yielding assets in the country. When dollars paid to buy productive capacity in Europe are then used to buy productive capacity in the US, it is, in the national accounts, equivalent to an exchange of ownership of assets. Or, when a US deficit on the current account is paid with dollars and the same amount of dollars is used to buy US debt on the primary or secondary market, the deficit is, in effect, being financed with US debt.

Hence, by relieving countries' monetary authorities of the obligation they felt to hold foreign exchange as redundant reserves, the evolution of international financial markets reduced the seignorage of the US. It arises now mainly from the various outflows on the current, capital and financial accounts of the US balance of payments less the acquisitions by non-residents of US assets, for payments by the country for other purposes are not reduced. One manifestation is the growth of international reserves held as dollars between 2000 and 2018 by \$ 5,543 billion.

When Kenen, Kindleberger, McKinnon and others put their arguments forward, they did not have to explain how trade and current account deficits could be incurred passively. Since then the trade and current accounts of the US have been in deficit from the end of the 1970s on, especially after the year 2000, when these deficits reached levels that nobody had thought possible. As the deficit on the US trade account grew from \$ 196 billion in 1997 to \$ 821 billion in 2007 and on the current account from \$ 141 billion to \$ 711 billion the question naturally arose as to how much more of US dollar liabilities the world would accept, its sustainability. A conference was held in 2005 at the Brookings Institution to discuss it. The bit of the conference concerned with the questions of passivity and exorbitant privilege is the hypothesis of Dooley and Garber, '... that the United States is passive and the foreign official sector is the active player in global imbalances.'²⁰² For that Dooley and Garber take up foreign investment; not US investment

²⁰¹ Kindleberger, *Europe and the Dollar*, 9.

²⁰² Dooley and Garber, 'Is It 1958 or 1968? Three Notes on the Longevity of the Revived Bretton Woods System'.

abroad, but foreign investment in the US. According to them countries like China want investments in the US, whether or not they are remunerative, because they are collateral guaranteeing US investments in these countries against expropriation, and, to acquire the means to invest in the US, these countries have deliberately brought about the current account deficits. By this argument sustainability is assured.

As both Eichengreen and Frankel state in their comments at the conference, the evidence is against the hypothesis; the US only accounts for about 10 per cent of foreign investment in China and other countries attract US investment without much investment in the US. They also point out that there never has been mention of such a motive for investing in the US or elsewhere; if there had been such a motive, there would have been some awareness of it instead of leaving it to be discovered in 2005. But the point that matters here is the supposed passivity of the US regarding its current account, which Dooley and Garber try to demonstrate by assuming that when the interest rate is higher the US demand for funds, i.e. investment less saving, or the current account, is lower and the private sector supply is higher. China's officials supposedly do not mind what the interest rate is and provide funds to lower the rate below what it would otherwise be to have a greater US current account deficit, more capital in the US and greater guarantees for US investment in China. Apart from doubts about the motive, this is not an explanation since it takes the relationship of the interest rate and the current account as given. Passivity here means that the relationship is extraneously given and permanent, therefore, to be accepted as such by the rest of the world. Obviously that part of the excess of expenditure over production that is the government budget deficit cannot be considered passive and, other things being equal, it adds to the current account deficit by its full amount. Dooley and Garber ignore this and dismiss the budget deficit as not being unusually big relative to GDP, which is irrelevant. Further, the term passivity means that the propensities to save of US households and firms as functions of the interest rate have to be taken as given and unalterable, though China and other countries are assumed to manage theirs.

Exorbitant privilege has been given another meaning; the higher rates of return US investments supposedly obtain outside the US as compared to foreign investments in the US.²⁰³ If the yields on specific types of investment do not differ much between countries, such higher rates of return are natural because direct investment normally fetches higher yields than portfolio investment, i.e. fixed interest securities of firms and governments and equity

²⁰³ Gourinchas and Rey, *From World Banker to World Venture Capitalist*; Habib, "Excess Returns on Net Foreign Assets."

from the secondary market, and the proportion of direct to portfolio investments is greater in US investments abroad than in foreign investments in the US. This difference of proportions, too, is as to be expected, for the rest of the world as a whole, especially Europe and Japan in their times of fast growth after the 1950s, provided more opportunities for attractive returns on direct investment than the US alone, though the biggest economy, so that the proportion of dollars held as portfolio investments by residents of countries other than the US has been correspondingly greater.

But Gourinchas and Rey argue that, not only are direct investments and equity larger parts of their foreign assets, but they fetch higher yields, as much as 3.3 percentage points higher for total gross assets from 1973 to 2004.²⁰⁴ US residents show greater tolerance of risk, i.e. they are “venture capitalists”. *A priori* this is unlikely; in theory when there is no premium on risk the returns are the same over the long run and a perpetual premium yielding average returns so much higher needs explaining, itself, as does the reluctance or inability of the rest of the world to profit from it. Their rates of return are calculated from “... the cumulated (*sic*) value of capital gains and exchange rate adjustments omitted from the current account measure.”²⁰⁵ As Curcuru et al. point out, this is a parlous procedure since the values of foreign assets are those the Bureau of Economic Analysis obtains from surveys, a different source to those of the current account.²⁰⁶ Their more direct estimates of rates of return give much the same returns for the various types of US investments abroad and the same types of foreign investments in the US. Gourinchas and Rey seem also to neglect FDI on the capital and financial accounts of the balance of payments. In the technical appendix they give an equation for the increase of the value of foreign assets that includes capital and financial accounts, but the text omits these accounts and is inconsistent with the appendix. Habib follows much the same procedure as Gourinchas and Rey to make the same calculations for several countries and does include FDI in the equation for the increase of foreign assets. But he combines it with the error term of the balance of payments and asserts that it is negligible, which it is not in the IMF’s International Financial Statistics Yearbook 1999 (line 78bdd); it rose from over \$ 100 billion in 1997 to just short of \$ 400 billion in 2007.²⁰⁷

²⁰⁴ Gourinchas and Rey, *From World Banker to World Venture Capitalist*, 11.

²⁰⁵ Gourinchas and Rey, 5.

²⁰⁶ Curcuru, Dvorak, and Warnock, “Cross-Border Returns Differentials,” 1499.

²⁰⁷ Habib, “Excess Returns on Net Foreign Assets.”

Current Accounts, Foreign Investment and Central Banks

²⁰⁸The Bretton Woods system was one that had been designed and agreed and its eventual replacement, the present system, is one that has not been explicitly agreed. Exchange rates are determined by the market, with some management, and the main international reserve and means of payment is the US dollar, though other hard currencies are also used to a lesser extent. Virtually all international payments go through US controlled channels. Gold has been demonetised and the SDR, which was supposed to replace it and be, at least, the main means of balance of payments settlement between central banks, is now only an appendage.

Most of this may have been foreseen by Triffin and others, but there have been a couple of phenomena that were not foreseen. One is that gold has fetched a consistently higher price after having been demonetised than before.²⁰⁹ Its price in 1978 exceeded \$ 800 an ounce, despite the sales by the IMF in 1975, after which it fell to fluctuate most of the time around \$ 200 – 400. Then, because of the crisis that began in 2007, it rose to over \$ 1,800 an ounce to fall again to \$ 1,200 -1,500. Most monetary authorities have been reluctant to sell their gold at the high prices and keep it as a reserve asset. Perhaps some would prefer SDRs, but not enough have been created to provide for the growth of international payments needs and the creation of more would require the agreement of the US and the main European countries, which is unlikely. The US is content with the present international financial system and in some European countries the authorities think creating more SDRs would be inflationary. Consequently, gold is still regarded by many as likely to remain a store of value if the international financial system has a crisis, certainly more likely than many currencies.

Equally unforeseen was that there seems to be no limit to the amount of dollars the world outside the US has been willing to hold or to deficits on the current account and outflows of direct and portfolio investment. Other countries with hard currencies, too, have repeated current account deficit. For instance, the current account of the UK has been in deficit uninterrupted since before 1996. Nevertheless, government officials and economists criticise the more recent deficits of the high wage countries less than they criticised the smaller ones of the time of the Bretton Woods system. Economists may have expressed unease over how much longer the world would continue to accept the outflow of dollars from the US, yet, since the Brookings Institution conference of 2005 to discuss the matter, there has been the financial crisis of 2007-8 and a repetition of the growth of the US current

²⁰⁸ {Citation}

²⁰⁹ An exception is Sraffa, who advised a young fellow of his college in Cambridge, Trinity College, to buy gold in 1964.

account deficits without the mounting stocks of dollars seeming to perturb the financial markets or the monetary authorities around the world.

The reason is that the international financial system allows holders of hard currencies to place that money in almost any country and obtain a yield. Much the greater part of almost any country's foreign exchange from exports and factor payments goes to firms and individuals directly or through the financial system, and it is to be expected that the bulk of it is used to acquire assets, service debt and avoid taxes. Taxes avoided are also a yield. Research by the IMF and the University of Copenhagen on the world stock of foreign direct investment has shown that 'Over a third of foreign 'investment' is multinationals cutting their tax bills'²¹⁰. 'Nearly 40 per cent of worldwide FDI "passes through empty corporate shells" ...' principally in Luxembourg, the Netherlands, Malta, Ireland, Switzerland and 'a number of British overseas territories and crown dependencies'. Apart from FDI, firms and individuals also have fortunes avoiding taxes and sometimes the law and hold them in jurisdictions where they are not reported. Rules that have come into force in recent years have eliminated much of the secrecy that protected such fortunes in banks in places like Switzerland and Jersey by imposing obligations to report bank accounts to the relevant tax jurisdictions, except that the US does not report such holdings or the income they yield to other countries.

In any but the smallest countries the supply of the total stock of the country's wealth, including productive capacity, urban and agricultural land, buildings, debt, equity, some infrastructure, art objects and so on, that is available to non-residents to buy is practically unlimited in comparison to the demand. An idea of the dollars generated by US deficits and available for investment is given by comparing the US current account deficits from the year 2000 to 2018, which totalled \$ 9,495 billion, from which should be deducted the increase of the world's dollar reserves over this time of \$ 5,543 billion to give \$ 3,952 billion. Part of the difference returns to the US as purchases of various types of existing assets by non-residents privately and by various state organisations and part as investment adding to production capacity. Even if all the difference were to be used to buy existing assets in the US, or even if the cumulative current account deficit was, the overall effect would be unnoticeable. Sometimes prices in some areas or of some assets rise, but usually not for long, and, as long as the restrictions the authorities in the US or any other country place on what foreigners may buy are kept to specific instances, firms and individuals wanting to invest for a financial return have ample choice. That part that goes to adding to production capacity is foreign saving adding to the stock of assets.

²¹⁰ Financial Times. September 9, 2019.

Current account surpluses imply that the country is acquiring foreign assets or reducing its liabilities to non-residents, including liabilities to non-residents denominated in the country's own currency, it being assumed that net payments for labour of non-residents and some other minor non-trade items of the current account can be ignored. If the country does not have a great accumulation of liabilities from the past, the surpluses become mainly an accumulation of foreign assets. A distinction must be made between countries with trade surpluses in which a large part of the export receipts go to the state and the others.

Among the former are a few countries with small populations that export oil or other minerals and the state owns the deposits. Oil exporters like Kuwait and Norway have set up sovereign investment funds to provide income when their oil revenues no longer suffice to meet their requirements, which, since their scope for adequately remunerative domestic investment is small compared to the amounts available, means investing mainly in other countries. In Singapore the state also gets more revenue than its authorities believe can be usefully invested domestically, but in this case from infrastructure and related services that are not depletable natural resources.

China is another special case. Its trade surpluses have been exceptional because of their magnitude, but the corresponding income has mostly gone to firms that may be private or may be owned or controlled by the state. It has surpluses to invest, but not a sovereign investment fund; the surpluses are invested by firms and state controlled organisations. That China's economic success has depended on the acquisition of technical knowledge from the high wage countries is now understood, the difference to before being that its firms and state have the means of buying that knowledge by buying the firms that have it. For the authorities of the high wage countries the purchase of their countries' firms' technical knowledge by Chinese firms jeopardises their future competitiveness as well as posing questions of security, and, so, they prevent sales as they judge necessary. It cannot be said that China should have had smaller trade surpluses by saving less or letting the domestic investment rate be higher. Its exceptionally high saving rate is like what had been observed earlier with other economies of which the income of the mass of the population was growing rapidly out of poverty, notably Japan and Taiwan, and, given the scale and pace of China's economic development, it is unlikely that more economically justified investments could have been found, especially not if some of the investments that did take place were, as has been said, wasteful.

After deduction of the state's surplus on its own consolidated accounts, current account surpluses are saving by firms and households. A country that, like Germany, has persistent surpluses on the current and trade accounts

and deficits or only small surpluses on the state's accounts, is continually transferring saving abroad. Its firms may be investing in creating or buying production capacity in other countries and individuals may be acquiring financial assets or immovable property or they may be adding to their bank deposits; in one way or another they are placing their money abroad. It can then be assumed that the assets acquired abroad are almost entirely acquired by people with higher incomes and more wealth than the majority of the population because they hold most of the shares in the firms, receive most of the profits and are paid more.

Saving transferred abroad to yield income from the assets they buy is the part of the country's saving that was not invested in adding to production capacity, infrastructure or R&D to increase income and improve competitiveness in the future. As the trade surpluses continue, the assets abroad accumulate from their reinvested yields and the trade surpluses assets, except when bad investments and economic crises cause losses. Before the First World War Britain benefited from its foreign investments by having trade deficits, a possibility that seems unavailable to Germany and other industrial economies of which the trade accounts have been continually in surplus. There is no prospect for these economies of being repaid in terms of goods since that would require that the main trade deficits, primarily those of the US, turn into surpluses, for which saving by households and firms would have to increase and the government to stop having deficits.

As counterpart, surplus countries would have to lower their saving rates and increase consumption. In Germany the authorities, the central bank in particular, have for long kept labour cost increases down on the grounds that they would otherwise be inflationary or would reduce competitiveness, although the European Central Bank (ECB) has been trying since the crisis of 2007-8 to increase inflation, with little success, and the trade surpluses have come close to 8 per cent of GDP. Higher wages so as to increase the proportion of GDP consumed would raise the demand for tradables and reduce the trade surpluses, which is not, in itself, inflationary, though the associated demand for untradables can be if there are hindrances to the supply.

Central banks do not consider the economic consequences of persistent trade or current account surpluses to be their concern. Their primary concern nowadays is to keep inflation near some desirable level, though there has recently been uncertainty as to whether that means an average or a ceiling. It used to be a ceiling before the crisis. Earlier on the duty of central banks had always been to keep the growth of debt and ensuing speculation under control, or, in the famous words of one Chairman of the Federal Reserve Board, McChesney Martin, 'to take away the punchbowl just as the party gets going.' It had become evident over more than two centuries that financial

markets do not regulate themselves and if credit is allowed to grow unrestrainedly it grows to the point that a collapse is certain.

If the earlier conception of their duties had remained in force, central banks would perhaps have seen the rapid increase of the current account deficits of the US from 2000 to 2007 as an indication of excessive debt within the economy. They would also have scrutinised the types of financing going on in the financial markets. In Europe more attention would have been paid to the domestic debt of Greece, Ireland, Spain and the UK. Because of the crisis European, especially German, investments in the US made losses, perhaps several times as much as was lost from loans to Greeks, though the ECB seems not to have tried to assess the amounts. However, as M. Trichet, the head of the ECB at the time of the crisis, remarked some time after his term ended, inflation had been at the desired level before the crisis and the ECB had done its duty.

McChesney Martin's simile for the duty of the central bank makes clear that control has to start early, for the later the bank acts the greater is the contraction that must follow. Once the party is in full swing the central banker who wants to take the punchbowl away must be willing to be unpopular. In the years just before 1929 the Federal Reserve could see a crisis was coming, but dared not act since it would have been blamed for the crash that was sure to occur and feared that, since it had only existed since 1913 and was new, its existence and powers would be put in question. It also seems to have been deterred from acting in 1996 by the reaction of the financial market to the observation about 'irrational exuberance' of the Chairman of the Federal Reserve Board and preferred to let things be. Stock markets did fall in the year 2000, but credit was allowed to continue expanding and the markets quickly recovered. If it had not taken another seven or so years to occur the crisis of 2007 would have been less severe, but, both then and in 1929, when it did occur the central banks were not overtly associated.

Central banks have not had to answer questions as to whether or not their objectives have been correctly chosen or as to why they did not see that conditions in the financial markets could result in a crisis. Instead they have been credited with having stopped the crisis and saved the euro, although the consequences for some countries, especially Greece, Ireland and Spain, could not have been much worse and popular hostility towards the euro has been growing. Central banks mostly acquired their autonomy in the decades before the crisis, yet it seems now more unquestionable than before. Before the 1980s the only two of importance that were autonomous were the US Federal Reserve system and the German Bundesbank, which was set up after the Second World War by the US in roughly the same form. Central bank

autonomy is now part of orthodox economic theory and is thought of as such by the political leaders and government officials around the world. It is required by the IMF of low wage countries, though compliance is often partial. With autonomy comes the freedom of central banks to choose their objectives and to be their own judges as to how well they have attained them.

To avoid similar financial crises in the future central banks are putting in place stricter prudential regulations, including capital requirements. The regulations are formulated by the Bank for International Settlements, a central bankers' organisation that by virtue of being international can propose standards common to all countries. Central banks demand of the banks under their authority that they examine how they would be affected by specific adverse circumstances, the so-called stress tests. All this seems to imply that judiciously chosen regulations and capital ratios are sufficient to prevent financial crises, which implies in turn that, had they been in place before, there would have been no crisis. If it turns out to be true that having the right prudential regulations and stress tests is enough to prevent crises, financial systems will have been made self-regulating and central banks will be able to ignore McChesney Martin's dictum to confine themselves to trying to regulate inflation.

4. THE EFFECTS OF CHANGES OF EXCHANGE RATES ON LOW WAGE COUNTRIES

The exchange rate is a macroeconomic quantity; a change has repercussions throughout the economy. A description of some of the repercussions in low wage countries is given here and, since the changes in these cases are nearly always falls of the currencies in terms of other currencies, the description is limited to devaluation. The repercussions can be put in three categories. Firstly, devaluation alters the distribution of income. Secondly, it alters the relations between financial stocks and flows. Thirdly, repeated devaluation gives rise to incentives that hamper economic development.

Income Distribution

The main effects of devaluation on the distribution of income occur through the price rises that follow. A broad description of how domestic prices rise has already been given; border prices of tradable goods rise straight away and induce rises in the prices of untradables when they enter into their production, directly or indirectly, as current inputs or capital goods, and, if profit margins rise, that increases the prices of those made with any capital goods. How fast final prices rise cannot be said *a priori*; presumably they rise faster the more efficient and competitive the market and the greater the devaluation.

As prices rise rates of pay that are fixed in nominal terms, mainly wages and salaries in the private sector, salaries of government employees and pensions not linked to the cost of living, fall in real terms. Categories of income directly linked to prices are profit from manufacturing, profit from wholesale and retail and non-wage farm income, but whether or not the total of each category increases in real terms depends on how production is affected. Profit margins rise when prices rise and nominal wages do not and that can have two opposing effects on the trade balance: increasing saving by shifting income to higher income groups and increasing investment. Higher profit margins accrue mostly to upper income groups, who usually have greater propensities to save, but they probably increase the incentive to invest if it seems that devaluation will be followed by stability. Other things being equal, the net effect is unlikely to be great.

Devaluation usually does have an effect on production. As pointed out earlier, by raising prices relative to wages it can bring unused capacity into production. If that makes the economy competitive enough, restraining domestic spending releases goods for export that would otherwise have been bought by residents, the trade balance improves, perhaps even with an increase of investment. But low wage countries do not normally devalue because of idle capacity but to improve the balance of trade, for which it is standard procedure to restrain expenditure to reduce the domestic purchase of goods and to prevent or slow inflation. Production is likely to fall, not rise. The countries may be producing their export goods to capacity whilst their balance of payments problems are the consequence of excessive imports, worse terms of trade, foreign debt and capital movements.

If the trade balance improves it is because the domestic purchases of the goods the country exports and imports are reduced as part of an overall restraint on domestic expenditure. That includes reductions of government spending, tax increases, restraint on lending by banks and higher interest rates. In particular, cutting investment by the government is an easy, reliable way of reducing expenditure. It is rare that a low wage country can noticeably improve its trade balance by buying less of its own export goods; Ghana cannot much reduce its own consumption of cocoa nor Bangladesh its consumption of garments. Hence, almost all of any improvement is the consequence of a fall of imports because of reduced consumption and investment. By itself this does not necessarily imply a fall of domestic production; for example, imports of machinery are reduced along with investment.

But the measures to restrain expenditure impinge on all the economy and, as a rule, are taken so far that production does fall. Normally their express purpose is to slow or prevent inflation on the grounds that inflation cancels the benefits of devaluation, or, in other words, reverses the drop of

the RER from the devaluation. Hence, they go further than needed for improving the trade balance and cause both unemployment and a fall of production. Demand is being reduced to counteract the tendency to higher prices from the rise of foreign prices in terms of the domestic currency, though the effects on the different categories of income directly linked to prices will vary. Overall, the consequences for domestic prices are likely to be slower, irregular increases according to the degree of competition within the economy and the severity of the measures, and, as that happens, if nominal demand does not increase it falls in real terms.

As prices rise, pensions, salaries of government employees, ranging from highly ranked administrators and judges to simple policemen and bus drivers, salaries of doctors and nurses employed in hospitals and so on all fall in real terms. Devaluation could be made less indiscriminate by raising, in step with prices, fixed incomes that do not enter directly into the costs of production of tradables. But economists and the IMF object to this for two reasons. One is that this lessens the restraint on domestic expenditure and the other is that they believe that markets should determine the structure of incomes, not the authorities. Against the second reason is the argument that devaluation decided by the authorities, too, is interference in the market and it has to be demonstrated that interference is incompatible with improvement. If, instead of a discrete devaluation, the currency depreciates because the exchange rate has been left to the market, believers in the reliability of the market will believe that the new exchange rate is in some sense a correct exchange rate, e.g. an EER. It is unnecessary to repeat here the arguments given earlier about such exchange rates.

Financial Stocks and Flows: External debt

In a developing country three types of financial stocks are affected by devaluation: external debt, amortisation of capital equipment and savings.

Devaluation raises the cost, in terms of the domestic currency, of debt denominated in foreign currencies and, even if the volume of domestic sales does not fall, only a rise in prices can prevent the liquidity and profits of non-financial businesses with such debts from declining. Well established, profitable firms, though they produce for the domestic market, are likely to be able to accommodate the effects of a single, moderate devaluation until prices have risen because their foreign currency debts are likely to be small, and, if they have problems, these firms usually have access to cash reserves or bank loans to tide them over until prices rise. New firms with foreign debt may not yet generate much, if any, profit, and are more likely to be bankrupted.

Financing obtained from banks in terms of the domestic currency can have the same effects as foreign currency loans if the banks, themselves, have foreign

currency debts. As a simple example, if a bank that takes a five year loan at five per cent and lends it at ten per cent, repayment in both cases being five equal instalments, it must raise its lending interest rate to $11\frac{1}{4}$ per cent to cover the interest cost of its own borrowing if the devaluation raises the price of foreign currency by 25 per cent. If the bank wishes to recover its margin of five per cent to cover its administrative costs, risk and profit, its lending rate is $16\frac{1}{4}$ per cent. Repayment of principal increases by 25 per cent.

Demand restraint intended to keep prices from rising compounds the effects of the immediate rise in the costs of external debt by a fall in domestic sales. If the devaluation is big, firms become unable to stay current on their loans, banks restrict credit because of the spreading of credit problems, and the demand restraint that was meant to restrain price rises, along with the rise of prices of imported inputs, turns the external financing into a source of widespread business failures.

Financial Stocks and Flows: Amortisation

Since developing countries as a rule import practically all their plant and machinery from developed countries, the amortisation for replacing their existing equipment and the savings for buying new equipment must be adjusted to the exchange rate. Amortisation is rarely, if ever, adjusted in this way, so that the amortisation funds a prudently run firm sets aside for the replacement of old equipment become insufficient if devaluation occurs.

Usually equipment is amortised at historic cost in terms of the domestic currency. Then the shortfall for replacing it is equal to the devaluation. Amortising at replacement cost may be difficult or impossible, even if the firm is eager to do it. One reason is that the firm would need to foresee the devaluation several years in advance. In a country that has not devalued for a long time the expectation of devaluation will arise only weeks or months before the event at the earliest, whereas the equipment may have been bought several years earlier. If a piece of equipment being amortised evenly over five years is four years old and devaluation raises its replacement cost by 25 per cent in terms of the domestic currency, its amortisation in its last year would need to be 45 per cent of its value. If the firm has not placed this amount in its cash reserves, it will have to borrow from a bank or issue new shares to replace its equipment.

A second reason that amortisation at replacement cost may be impossible is that the authorities regulate how amortisation may be calculated, both because it affects the taxes paid by the firm and because rules are needed to protect shareholders, creditors and the public. Hence, even if a firm wishes to amortise at replacement cost, it may be prevented by the rules and it will

not persuade the authorities that its forecast of a future devaluation should be used for calculating that cost.

Financial Stocks and Flows: Savings

Since devaluation causes inflation, it reduces the real value of savings in the form of cash, bank deposits, fixed interest securities and acquired pension rights, the principal financial savings of lower and middle income groups. If its purpose is to lower some components of the country's production costs of tradables in terms of foreign currencies, this reduction is a redistribution of wealth.

To some extent, it can be compensated by adjusting the nominal value of the savings in step with prices. This is easier with pensions in a state run system, but whether or not it would be financially sustainable is a separate question that can only be answered by assessing the future receipts and payments, as well as future exchange rates. In most countries that devalue frequently it seems that wage earners and middle income groups have been the losers in the redistribution of wealth consequent on diminution of the real values of pensions. Nevertheless, adjusting the nominal value of savings to compensate for higher prices may be impractical, especially in fully funded schemes, since it would raise the liabilities of the institutions holding them without corresponding gains in their assets, and would require higher contributions.

Alternatively, the real value of savings can be preserved by adjusting interest for inflation. In practice this happens rarely; the interest received by lower and middle income groups in countries that devalue often does not suffice to offset inflation. The economic difficulties that led to and followed from devaluation have their repercussions in the banking system, which protects itself, in effect, by reducing its liabilities to those who cannot negotiate their own terms for depositing with the banks.

Effects on economic incentives

A single, moderate devaluation that is not expected to be repeated in the foreseeable future is unlikely to alter economic behaviour. Firms and households adjust to the new exchange rate and continue as before. But repeated devaluation creates the expectation of more. The following lists a few of the more important consequences.

Holders of wealth make windfall gains in terms of the domestic currency by converting their financial assets into foreign exchange, i.e. capital flight. Exchange controls can limit the outflow, though they may not stop it altogether. This does not mean that they are useless or undesirable; they prevent the sudden large outflows that afflict countries that do not have them and, properly managed, keep them small. Yet, over time, the loss of international reserves from capital flight is greater than the gains expected

from devaluation. According to such estimates of flight capital as there are, the amounts that have accumulated often exceed the foreign debt of the countries from which they came.

If, as with some countries of Latin America, the accumulated flight capital is large, the owners, many of whom have business interests in their home countries, have an incentive to press for devaluation of their home country currency since they are usually well informed and able to remove their funds at the right moment. Mexico's budget deficits in the early 1990s were financed in great part by Mexican flight capital at high interest rates, usually over 20 per cent, and short term. This was the capital that moved out first when crisis began at the end of 1994.

Devaluation gives windfall profits to firms producing exports which then have an incentive to press the authorities for more devaluation as a source of profit, rather than to improve efficiency and seek new products and markets. This is probably why Pakistan's exports of cotton products have low unit values compared to those of competitors.²¹¹

Since amortisation does not normally suffice for replacement of equipment, firms resort to more bank loans. The greater the rate of devaluation, the more indebted they become.

Inflation and the expectation of inflation become permanent and become obstacles to the evolution of a sound financial sector. They deter holding savings in the form of bank deposits and fixed interest securities, which usually offer yields below the inflation rate. Long run savings, notably pension plans, lose most of their expected value unless they are indexed to inflation. When they are so indexed they depend on the state's budget and add to the political complications of expenditure restraint.

Trades unions have to press for repeated wage increases to keep pace with inflation, which the authorities try to oppose by making demand restraint more stringent. The results are higher unemployment and impoverishment of workers who do not belong to strong unions and of the middle classes. Depending on the political circumstances, the authorities can, over the long run, try to break the power of the unions, which then adds to the effect of devaluation in making the distribution of income more uneven.

²¹¹ World Bank., "Pakistan: Export Diversification and Trade Policy," 70.

CHAPTER 4

DOCTRINE AND CONSEQUENCE: PAKISTAN'S CASE

1. HISTORICAL: FIRST STAGES

This chapter describes the consequences for a country without abundant natural resources of having persisted in trying to follow its supposed comparative advantage and failing to acquire the ability to make complex goods. It is necessarily historical; Pakistan's economy as it is now and how it came to be as it is has to be understood as the outcome of decisions made in the 1950s and 1960s and conformity to orthodox doctrine since. It also concerns the present and the future; economic doctrine now is much the same as before and the main modifications of the orthodox economic advice the country has been receiving have been to deal with problems that have been its consequences, such as foreign and domestic debt, loss of sources of government revenue, capital flight, corruption and social tension. Apart from these, but beyond the scope of this work is the multitude of problems arising from the degradation of the environment, urbanisation and depletion of natural resources. If it is the policies of the last seven decades that have led to such poor results one choice is to examine these policies and the reasons for them and, using the arguments of the preceding chapters, to see what could have been brought about by other policies. The alternative is to continue as before and blame the country for the results.

Since there has always been much unemployed labour and wages always been low, conformity to comparative advantage has meant industries should be labour intensive, which means little capital per worker. From the start cotton and jute textiles seemed the obvious choices and the preceding chapters explain why, contrary to orthodox theory, Pakistan's income from these industries has been low in relation to the investment costs. The country produces and exports now much the same low wage goods as in the 1960s, which, combined with capital flight and the cost of servicing its foreign debt, has several times obliged it to reschedule its foreign debt. Change requires, among many things, improving education at all levels, reversing years of deterioration of the administrative system and changing the tax system. It is unrealistic to believe that any of this can be accomplished without more government spending, whereas the consequence of trade liberalisation has

been the reduction of government spending as a share of GDP to a third or a quarter of that of a high wage country because of a lowering of import duties, which had once accounted for most government revenue. Repeated devaluation has done little to improve the trade balance but has provided an incentive to transfer capital abroad and has ensured nominal wages have stayed low in terms of foreign currencies.

What follows is not meant to be a comprehensive account of Pakistan's economy but an account of how the pattern of foreign trade that was chosen explains much of the country's lack of development and to describe the repercussions within the economy to the extent they are part of the explanation. Its interest is that the experience of many low wage countries has been similar. They, too, produce the same low wage goods they have been producing for decades, depend on foreign aid, devalue repeatedly, reschedule their foreign debt and are short of government revenue. Their manifestations of the repercussions vary as their internal workings differ, but much of what they have in common will be recognised in the limited scope of this account.

The first steps

When British India was partitioned in 1947 into two independent countries, Pakistan and India, Pakistan consisted of East Pakistan, which became Bangladesh in 1971, and West Pakistan, the present Pakistan. Neither part of Pakistan had any modern industry. East Pakistan produced jute and West Pakistan cotton, and, since the mills to which the two crops had always been sent were now in India or the UK, the crops became the country's main exports.

In 1954, to stimulate investment in the production of cotton textiles, the government banned imports of textiles. Shortages resulted and the prices of cotton textiles rose so high for the next few years that new mills often recouped their investment costs within a year. Bank financing in rupees was ample, so the demand for equipment, along with other demand for imports, exceeded the foreign exchange available and foreign exchange had to be rationed. No stimulus of the same kind could be given for jute goods, for the demand for them was almost entirely foreign.

By 1956 Pakistan exported yarn and cloth, but foreign exchange became ever scarcer, to the point of causing shortages of goods deemed essential and preventing investment. In 1958 the government of General Ayub Khan, which had seized power that October, began a scheme of subsidies of exports of manufactures similar to the Exchange Settlement Certificates started in Taiwan in 1949²¹², the Export Bonus Scheme (EBS). Exporters of

²¹² Lin, *Industrialization in Taiwan: 1946 - 1972.*, 43.

specified manufactures received “bonus vouchers”, which were entitlements to buy foreign exchange at the official exchange rate and could be sold on the open market to import various goods not deemed essential, such as cars, air conditioners and radio sets. The “bonus rate”, the proportion of the entitlements to the foreign exchange from an export, was fixed by decree according to the type of export, usually 30–40 per cent of the rupee value, and the price, the “premium” was usually 150–180 per cent. Hence, when the bonus rate was 30 per cent and the premium 150 per cent, 45 per cent more rupees were received from the export than at the official rate. Goods considered essential continued to be imported at the official exchange rate.

As an emergency measure, the scheme was successful and resulted in a sudden increase in exports, but, against the advice of the German adviser who had proposed it, it was made permanent. Its appeal to the government was that it increased the profitability of manufacturing for export of all goods to which it applied and, therefore, stimulated investment in industries that did not have a domestic market that could be protected, notably of jute. Over the next ten years industrial output and exports of manufactures rose fast (industry averaged growth of 9.0 per cent and total exports 13.3 per cent p.a.). GDP grew at 6.9 per cent p.a.. In the opinion of some foreign advisers to the government, Pakistan was a successful economy, about to “take off” and had the beginnings of a capital goods sector. Most economists compared its economic performance favourably with India’s.

Yet there were signs that the economic growth was not what it seemed. One, so obvious and familiar as to seem normal was that foreign exchange remained scarce in relation, not only to demand, but also to consumption needs. Since exports were growing fast and foreign aid inflows were high, the scarcity of foreign exchange should have been reduced, if not eliminated. Instead, the ratio of debt service to exports increased, the premium on bonus vouchers did not decline and the rationing did not become less strict. Despite the investment in manufacturing and in increasing the production of cotton and jute, Pakistan’s exports in terms of current US dollars in the 1960s remained less than they had been at the time of the Korean Boom in 1953.

A second sign was that the mass of the population was not noticeably better off. To some extent this was the result of deliberate policy. The government’s “strategy” was to raise the saving rate by concentrating income, ‘to shelve for the distant future all ideas of equitable distribution and welfare state’²¹³ and to keep increases in social expenditure for the poor, like education and health, to a minimum. In the minds of the government’s economic advisers, ‘inequalities in income contributes (*sic*) to the growth of

²¹³ Haq, *The Strategy of Economic Planning: A Case Study of Pakistan.*, 30. Similar statements can be found in the Second Five Year Plan.

the economy, which makes possible a real improvement for the lower-income groups' and the 'concentration of income in industry facilitates the high savings which finance development'²¹⁴.

A third was the lack of interest on the part of the owners of the new industries in the production of capital goods or any technically more complex industries. Such machinery as was produced in the country on an industrial scale was simple items like pumps needed for agriculture and fans for houses. Some looms were produced in workshops for the domestic textile industry, but were copies of imported models with the drawbacks to be expected of old design, low precision and poor materials. Pakistan depended totally on imports for the plant and machinery of industry and for nearly all other engineered goods and chemicals.

2. VALUE ADDED AT WORLD PRICES

Protection and value added

Part of the explanation of the foreign exchange shortages first became apparent in 1965 from the calculations of effective protection of several industries by Soligo and Stern, according to which the value added at domestic prices was greater than the value added at world prices in every case.²¹⁵ They implied that the income these industries generated when measured at domestic prices was not matched by the net receipts or saving of foreign exchange. These were not results peculiar to Pakistan; calculations in other countries for various industries gave similar results. Lewis and Guisinger later carried out the same kind of calculations for Pakistan, with the improvement of using observed import and export prices in place of the prices Soligo and Stern had inferred from tariffs, and obtained similar overall results.²¹⁶

According to the proponents of ERP or DRC, such results show inefficiency; if the industries had been efficient, they would not have needed protection and, in those cases where there was more protection than necessary, either competition would have brought the prices of the goods below the levels allowed by the protection or there would have been oligopolies, in either case easy to spot.

As shown in Chapter 3, these assertions of inefficiency do not follow from such comparisons of value added, for they neither take account of the cost of the investment involved nor use the prices of producers in high wage countries, where markets were competitive and production presumably efficient. Pakistan exported cotton and jute manufactures and the high wage countries

²¹⁴ Papanek, *Pakistan's Development: Social Goals and Private Incentives*, 242.

²¹⁵ Soligo and Stern, "Tariff Protection, Import Substitution and Investment Efficiency."

²¹⁶ Lewis and Guisinger, *Measuring Protection in a Developing Country: The Case of Pakistan*.

imposed a variety of tariffs and import restrictions on them. Other low wage countries' exports were similarly hindered by the protection of the high wage countries. If the ERP calculations had allowed for the effects of this protection and had used the prices of the high wage countries as world prices, instead of the prices the exporters received, their results would have been different.

Inefficiency there may have been, but not enough that the high wage countries did not have to protect their own producers. No attempt was made by the proponents of ERP or DRC to show inefficiency in the case of Pakistan with direct evidence from factory data. It would have been simple for cotton and jute manufacturing; it only needed comparison between countries of current inputs per unit of output. For cotton textiles inefficiency would have been mainly too much cotton per unit of yarn or too much yarn per unit of cloth. There is no evidence that Pakistani textile mills were wasteful in this sense. The coefficients derived from the Census of Manufacturing Industries (CMI) and used to obtain the rates of return referred to below seem to have been normal.

What was inefficient was to rely mainly on investment in cotton and jute textiles for obtaining more foreign exchange when the high wage countries were protecting their own industries. As usual, selling to unwilling buyers fetched low prices. The consequence could be seen from the calculations for cotton cloth by the method described in Chapter 3 assuming a zero wage. The coefficients for production and investment were derived from the CMI and the data of individual firms and the prices were import and export unit values. For the years 1954, 1956 and 1959/60, the rates of return years were 14.8 per cent, 7.5 per cent and 11.7 per cent. These are measures of value added as a ratio of capital cost, i.e. output:capital ratios. Positive wages would have given lower figures. As output:capital ratios these figures are what would be expected of capital intensive industries, whereas textiles were assumed to be labour intensive. Confirmation that this was not the result of inefficiency comes from calculation of the rate of return for the same product, using the same method and the same coefficients, with the input and output prices of a high wage country. With the UK prices of 1959 the result was 48.3, which is what is expected of a labour intensive industry.

In other words, the investment in textile manufacturing was not as productive as was thought; the foreign exchange gained or saved from exporting cotton and jute manufactures, rather than the raw cotton and jute, was low compared to the foreign exchange cost of the investment and less than it appeared to be from the domestic prices. Since the wage was set at zero in these calculations, the rates of return were purely ratios of foreign exchange gained to foreign exchange invested and, since virtually all Pakistan's plant and machinery was imported, they were the equivalent to Pakistan depending on a capital goods sector of low productivity. The lower the rate of return, the less new investment or consumption a given investment could yield.

That the calculations are confined to early years does not mitigate their implications for the present. Firstly, the protection of the high wage countries increased in the following two decades and similar calculations in later years would almost certainly have yielded even lower rates of return. It was partly removed with the end of the Multi-Fibre Agreement in 2005, by when textiles had long been a low wage product. It is improbable that the Agreement would have been ended if they had not. While it lasted, the EBS made foreign exchange receipts seem greater than they were to the authorities and their advisers, who looked only at the growth figures, not at the net foreign exchange receipts allowing for the foreign exchange cost of the cotton and jute used in manufacture. It also inflated the profitability of investment, as was its purpose, and caused the demand for investment to exceed what the available foreign exchange permitted. Nevertheless, the cotton textile industry was efficient enough that, after the EBS was ended in 1973, its exports continued much as before. But Pakistan has remained a country exporting only low wage goods.

Secondly, Pakistan became dependent on foreign aid, which it would not have become, or to a lesser degree, had its cotton and jute manufactures fetched the same prices as the equivalent products in the high wage countries. In 1961-62 external debt service reached 27 per cent of exports and fluctuated above 20 per cent until 1998, descending slightly below 20 per cent in only three years but sometimes exceeding 30 per cent.²¹⁷ As a comparison, since these manufactures accounted for 73 per cent of total exports in 1970, debt service probably exceeded the value added, i.e. the direct foreign exchange gain from investing in the manufacturing equipment. In 2017-18 it was equal to 23 per cent of exports.

Economists from high and low wage countries deplored the protectionism of the high wage countries, but their concern was the quantity of exports, not the value added. They tried in various ways to estimate how much more the low wage countries might have exported with freer trade without paying attention to the effects that protection had on value added. So, they continued both to advocate cotton and jute manufacturing as labour intensive industries suited to the factor endowments of low wage countries and to assert that the value added was low because of the inefficiency of the producers.

They were constrained in two ways. One was the belief common among economists in the 1950s and early 1960s that the economic backwardness of the low wage countries was to be explained by the lack of "entrepreneurship". It is now out of date, but the belief seemed obvious at the time for countries that, like Pakistan, had virtually no industry and had to start economic development by fostering entrepreneurship, which is what

²¹⁷ Government of Pakistan, *Pakistan Economic Survey 1999-2000*, Table 9.3.

the ban on textile imports and the EBS did. Examining the costs and economic returns was at this stage beside the point. But, judging by the demand for investment and the premium on bonus vouchers, there was no dearth of entrepreneurship. It was not surprising, for there were a number of Muslim communities with long commercial traditions, sometimes with international networks, who immigrated from India and in a few cases from Africa, and there were some Pakistani by origin. The absence of industry was just a consequence of how industry was concentrated in India before 1947 and was true for many other parts of the subcontinent.

Economists were also constrained by their notion of comparative advantage and the depiction by the Haberler representation of the effects of tariffs. Since the representation depicts countries as producing at their production frontiers and all tradables are assumed to have world prices, it can only depict the effect of a tariff as a change of position on the production frontier; the ratio of value added to fixed investment is outside its scope. Hence, if the private firms of a low wage country could not obtain the same income from an activity as those of the high wage country, which was assumed to be efficient because it was a market economy, the reason had to be some inefficiency of the country, which, given the presumed efficiency of competition in the private sector, could only be the effect of the low wage country's protection.

The appearance of economic success ended in late 1968 with a popular uprising against the Ayub Khan regime. Having fostered income inequality and discriminated against East Pakistan since it took power, the regime could offer nothing to allay the discontent that was manifest in both wings of the country and was overthrown in a few months. Elections were held in 1970. In East Pakistan the embitterment of the population was expressed in a vote that ultimately led to its separation at the end of 1971 to become Bangladesh. What had been West Pakistan became what is now Pakistan.

3. GLOBALISATION AND THE ILL-ADAPTED ECONOMY

The Economic Legacy

But the form of Pakistan's economy had been set; most of the change since has been of scale rather than of kind. The main industries, apart from cotton products, produced simple goods for domestic consumption, among them food processing, cement, simple machinery like fans and water pumps, and miscellaneous items like light bulbs and toiletries. Assembly of motor vehicles with the protection of high import duties started in the 1950s, though there have been several changes of brands since then and the locally made component has increased. Otherwise the closest to complex goods

production have been nitrogenous fertilisers from local natural gas and refining of petroleum by firms controlled or in partnership with foreign firms. Production of capital and intermediate goods consisted of simple machinery, mainly water pumps for agriculture, cement and some basic steel and iron construction material. Since then the main new industrial activities have been the steel mill, which was completed in 1985, and some weapons. Agriculture has diversified to provide some products that used to be brought in from East Pakistan and some that were not consumed in Pakistan before. Parts of the services sector have been changed by technical changes occurring around the world, as with communications, information and medical care, but other parts have evolved little.

Exports illustrate the lack of change; cotton manufactures account for around 50 per cent of exports of goods and the same goods that accounted for 78.5 per cent of exports in 1973-74 accounted for 76.8 per cent in 2018-19. Some 15-20 per cent are rice, leather goods, synthetic textiles, surgical instruments and sports goods. Since 1974 some agricultural products have been added, namely fruits, sugar, wheat, meat and meat preparations.²¹⁸ Pakistan exports no complex goods made in the country and the unit values of its main exports, in particular those of 70 per cent of its textile and clothing, are low in comparison with its competitors'²¹⁹.

It is this inability to change the economy and the consequences for the people that the following attempts to explain. In other parts of the world change in the form of what is now termed globalisation became evident in the 1970s, but Pakistan's economy had no part in it. Japan was increasing its exports of complex goods fast and, along with other high wage countries, was transferring production to Korea and Taiwan, economies that were also growing rapidly. Production was also being transferred by high wage countries to other East Asian economies, notably Malaysia and Thailand, mainly through subcontracting simple production and direct investment. A new form of bank financing for the low wage countries and a new need for it came from the increase of revenues of the oil exporting countries after the increases of oil prices in 1973 and 1979, what was termed recycling petrodollars. For Pakistan the main change was the employment generated by the Arab oil exporting countries. Remittances by Pakistanis abroad grew so fast that net factor income from abroad, which had been negligible in 1973, reached 8 per cent of GDP in 1980-81 and is now almost equal to the receipts from exports of goods. Countries like Bangladesh, Egypt, India, Philippines and the Sudan, have had movements of workers and flows of remittances of a similar scale.

²¹⁸ Government of Pakistan, *Pakistan Economic Survey 2018-19*; Government of Pakistan, *Pakistan Economic Survey 1980-81*.

²¹⁹ World Bank., "Pakistan: Export Diversification and Trade Policy," 70.

In time globalisation led to greater insistence by the western high wage countries on free markets. By the end of the 1970s the high wage countries, especially the US, and institutions like the World Bank had begun to demand the reduction and eventual removal of protection of domestic production and of various practices for increasing exports. Fixed exchange rates were also being replaced by crawling pegs or floating exchange rates in accordance with the doctrine that the market should determine exchange rates. The high wage countries had removed obstacles to their own international movements of private capital, though there was less insistence on opening the capital accounts of low wage countries. But the high wage countries' own protection against the exports by low wage countries of textiles and other manufactures exported by low wage countries grew stricter; the Generalised System of Preferences covered almost all manufactured and agricultural exports of the low wage countries, except minerals. In particular Pakistan's textile exports were constrained by the MFA, which continued until 2005, and both the US and the European Union still levy tariffs of 10 per cent on them.

Exporting manufactures to import complex goods

The specific importance of the textile industry is that it was to be the main means of adding to the availability of foreign exchange and foreign exchange had no substitute for obtaining complex goods, in particular capital goods, or servicing foreign debt. Leaving the debt aside, complex goods are, in Pakistan, as in any other low wage country, used in all parts of the economy, except, perhaps, in some backward, stagnant activities in areas that interact little with the rest of the economy. Virtually the whole economy, therefore, depends directly and indirectly on foreign exchange.

This is evident in Pakistan's agriculture, where production depends directly on fertilisers, pesticides, tractors, vehicles and communications, all of which are either complex goods not made in the country or are made with imported complex goods. Such causes of output increases as soil conservation and extension of the irrigation system, which do not use imported goods directly, were made possible by investments beginning in the nineteenth century and continuing to the present in canals, earthworks, barrages and dams, as well as investment in reducing the harmful side-effects of the canal system of water-logging and salinity and in exploiting subterranean aquifers. But as long as Pakistan has existed imported machinery has been used for the creation and upkeep of such works.

Many services, too, depend on complex goods. Transport, communications and medical services are obvious. Others are the retail of complex goods, like television sets, computers, vehicles, air conditioners and washing machines. There are also the services for the repair and maintenance of these goods, ranging from roadside mechanics to the service facilities of

major foreign firms. Office activities usually require communications, depend more and more on computers and are sometimes concerned with the import, sale and insurance of complex goods.

What distinguishes manufacturing is that capacity is increased directly by investment, which is not true of agriculture or all services. Neither in Pakistan nor elsewhere is the relation between agricultural capacity and investment a direct one. In Pakistan agriculture has been increased more by the supplies of water and soil and by farming practices, particularly the varieties of seed, than by direct investment or foreign exchange. Much of the investment in agriculture has been remedial, reducing the waterlogging and salinity resulting from the age of the canals, and not addition to capacity. That output has been growing and can be expected to grow more, and that several products, including meat, fruits and rice, are exported are more attributable to better practices and seeds than to machinery. In services, other than transport, communications and medical care, capacity often does not depend on equipment consisting of complex goods. Most retail and much of repair and maintenance, education, health care, personal household services, hotels and restaurants, among other things, cater to consumption and their supply is usually not restricted by the availability of complex goods. Data processing and technical services for foreign clients require on the job training and investment in the education system and, though education may be an investment in the economic sense, the link between it and these services is not direct. Extraction of minerals is ignored here because there is little worth extracting in Pakistan.

Hence, the only apparently reliable way of increasing the availability of foreign exchange by domestic production, at least in Pakistan, is by investment in manufacturing. There may be some forms of manufacture for which little or no investment is needed, assembly of garments for example. But, since these are activities not limited by the investment in capital equipment and in which many low wage countries compete, they depend on keeping nominal wages down in terms of foreign exchange and for individual countries their scope for growth in the long run is uncertain. An alternative to manufacturing is to provide data processing and technical services from within the country to clients abroad, which some countries, like India and the Philippines, do. Often these services are simple, low wage activities, though, since even the simplest of these services require literacy and training, the pay of those providing them is above that of untrained workers, more like that of better paid clerical workers. When they are technically advanced services they are not low wage activities, but are provided by people with suitable training, who will then have obtained their training in the country's institutions of higher education. Pakistani firms mainly provide the simpler kinds of these services. Providing technically

advanced services, which are equivalent to complex goods, is limited by both the education system and the need for proprietary knowledge.

Saving and foreign exchange with fixed exchange rates

Greater inequality of income and neglect of the social sectors, such as education and health were supposed to increase the saving rate, but, as long as exchange rates were fixed, up to 1982, saving available to the economy was mainly determined by the supply of foreign exchange, which was always exceeded by the demand. Pakistan devalued twice in this time, in 1956 and in 1972, the second time from Rs. 4.76 to the dollar to Rs. 9.61 to put an end to the EBS. The authorities, who controlled all sales of foreign exchange, wanted to direct investment to activities they favoured after making sure of the supply of necessary imports, such as medicines and staples and allowing some imports of other consumer goods for which there was demand. Hence, the foreign exchange controls were detailed and had to be accompanied with control over domestic bank financing for investment. Annual plans were prepared for the economy as a whole with, among other things, estimates of what the receipts and various uses of foreign exchange would be and set out how the foreign exchange and bank financing were to be allocated, with adjustments as needed as circumstances changed. Of the uses some, such as debt service, representation abroad and changes of reserves, had to be decided on before imports.

A consequence was that saving, as given in the national accounts, was mainly determined by the decisions of the authorities on private and public sector investment, of which the bulk was constrained by the availability of foreign exchange. The estimate of saving in the national accounts was calculated as the sum of investment and the trade balance and, since the trade figures were, in principle, available from the customs and central bank data and since public sector investment was, in principle, already known, all that was needed was an estimate of private investment. Large scale gross fixed capital formation accounted for the bulk of investment and the data were readily available for the national accounts. Investment by the larger private firms, particularly in manufacturing, could be calculated from the accounts of the firms or from their reports to the census of manufacturing industries, but, since these investments included imported capital goods and the authorities had to approve the foreign exchange for them as well as the bank domestic credit, most of the needed data were there. Other forms of private investment, such as changes of inventory, investment in agriculture, by small firms or workshops and by households, were overall a small part of total investment and had to be estimated by a variety of methods.

In theory it can be argued that public sector investment could have used less foreign exchange if the projects had been suitably chosen and more

labour had been used in place of equipment. Perhaps there could have been some such reduction of foreign exchange use, but it would have been minor. In practice labour was usually used instead of imported machinery when the costs and time allowed it, for projects for which there was financing by agencies like the World Bank were evaluated and carried out using economic criteria that specifically allowed for the abundance of unskilled labour and the scarcity of foreign exchange. And usually there was little choice. Much public sector investment consisted of infrastructure, which required stocks of imported capital goods like earth moving and transport equipment, which lasted for several years and was used for a succession of tasks. Big projects, such as dams and some improvements of the irrigation system, could take several years, with a great part of the financing as foreign aid, and included, in addition to the earth moving and transport equipment, things like sluice gates and turbines, all of which were imported for the specific project.

Had more foreign exchange been available from reserves or a foreign loan or grant and had it been allocated to private investment the saving rate would have been higher. As long as the EBS and protection assured high rates of profit, there was no immediate limit to the demand for these imports. Saving was not constrained by the domestic propensity to save, for the banks were allowed to provide loans for investment when the foreign exchange was available. Little capital was raised from the public through the issue of shares or bonds and even less from the savings of those who controlled the investment. The only effective constraint on private sector imports of capital goods were the controls over foreign exchange. Hence, more investment meant more saving, though, if some of the additional foreign exchange came from foreign sources, that part of the addition to saving was foreign in origin.

If that additional foreign exchange had been allocated for private consumption, saving would have been less. Consumers would have bought more imports and would have spent more. In practice they spent mainly from their own means. They would, therefore, have kept less income as financial assets like bank deposits and government bonds, though they might also have borrowed from banks to buy items like cars. Putting it in reverse, less foreign exchange for consumption goods obliged consumers to spend more on items that had little or no foreign exchange components, notably housing, or keep more of their income as financial assets, i.e. as savings.

Hence, if concentrating income could have caused an increase of the saving rate as given in the national accounts, it was frustrated by the shortage of foreign exchange. According to the official figures, the gross domestic saving rate was 6.0 per cent of GDP in 1960/61 and varied between 10 and 15 per cent from 1962/63 – 1973/74, though it fell to -0.6 per cent in 1966/67 because of war in 1965. At its highest the saving rate was comparable to the

saving rates of the high wage countries, though low compared to the saving rates of some of the East Asian economies. Investment (gross fixed capital formation and inventories) was constrained to 18–19 per cent of GDP, since the amount by which it could exceed saving was limited by the external financing available, in those days foreign aid.

Concentrating income may all the same have increased the saving rate in reality, but to the benefit of other countries. Saving was higher than the national accounts' figures, with the difference going abroad undetected. This part of saving has to be taken as given, whilst the saving available to the economy is that given by the national accounts. Calculating saving from its identity with the sum of the trade balance and investment gave underestimates because over-invoicing imports and under-invoicing exports, which were known to be common and an invisible outflow of income, made the trade deficit seem bigger than it was. False invoicing like this was common to many low wage countries and a reaction to awareness that foreign exchange would remain scarce. Pakistan's authorities could probably have restrained such practices if the political leaders had been less indulgent.

Globalisation Sets the Trade Rules

Changes of regime

A new government came to power at the end of 1972 and began, as it had declared it would, by nationalising the biggest industrial firms, the banks and the shipping companies and by abolishing the EBS. The first steps were taken to set up an organisation for managing the nationalised industries, whose efficiency seems to have improved at first²²⁰, though that of the banks and the new national shipping company did not. It also planned and started construction on a new deep water port, which began operations in 1980, and the steel mill. New universities were created, in particular for towns where there were none.

But in 1973 there followed the nationalisation of rice mills, vegetable oil mills and primary schools, excepting those run by foreigners and religious organisations, Christian missionaries in particular, to which the upper income families sent their children. Since they occurred after assurances that there would be no more nationalisation, there was little prospect of attracting private investment in industry, especially to produce complex goods. This round of nationalisation was undertaken mainly to favour farmers and the rural population, for the government had come to be dominated by agricultural interests and its head, Mr. Bhutto, came from a wealthy land owning family. In particular, the nationalisation of schools was an

²²⁰ Zaidi, *Issues in Pakistan's Economy*, 156. Zaidi quotes the World Bank.

unsuccessful attempt to transfer teachers from the towns to the villages, where schools were scarcer, and harmed, rather than helped, education for having been carried out without preparation, entailing costs for which the budget had no additional revenue and ruining many small private schools in cities. Nothing was done to raise the standards of education in schools or of science and engineering in the existing or new universities and colleges. Government expenditure on education remained under two per cent of GDP.

Another uprising in 1977 led to the replacement of this government by the military regime of General Zia ul Haq, which favoured the private sector and began to reverse the nationalisations. From then on all governments have adhered, more or less, to the economic doctrines of the international multilateral institutions, especially the World Bank and IMF, and have taken steps to make markets for goods, capital and labour more competitive and to attract foreign investment. Some big infrastructure projects that were carried out did not necessarily have the approval of these institutions, in particular the new port of Gwadar. Education became a rapidly growing business and private profit making primary and secondary schools have proliferated. So have institutions of higher education of which the degrees are recognised by the provincial or federal government. But the standards of primary and secondary education in the public schools seem not to have risen.

The rules of trade

It was not obvious then, but the early 1970s were the last years that Pakistan could increase the protection of its industries with trade barriers or use the various devices that other economies had been using but incompatible with the tenets of free trade, like subsidies, performance targets and state help in getting proprietary knowledge on favourable terms. During the 1980s the high wage countries, particularly the US and Western Europe, became less tolerant of such practices, which they maintained were economically inefficient for the low wage countries and unfair toward themselves. They and international organisations, especially the World Bank, the IMF and the GATT/WTO, began to insist on greater conformity to free trade. A country that needed financing from the World Bank or IMF, as did Pakistan, was obliged to conform more to their doctrine that the proper functions of the state did not include helping private firms or running activities that could be run by the private sector and that private firms were to be left to compete with an appropriate regulatory framework but little protection. Even when the financing did not explicitly require liberalising trade, it ruled out contrary changes.

Trade negotiations under the auspices of the GATT/WTO gradually obligated low wage countries to lower their tariffs, remove quantitative restrictions on imports and stop using devices that were considered to be

inconsistent with the ideal of free trade. Pakistan's customs duties were eventually reduced to at most 25 per cent, the majority being 5–20 per cent, which economists in the international organisations judged to give the right degree of protection for infant industries, not so high as to permit gross inefficiency but high enough to protect infant industries that could become competitive. Complete motor vehicles and parts, which had duties of 35–90 per cent, were the exceptions.

India was better able to delay or avoid lowering the protection its industries needed for two reasons. One was that these industries were big enough that the authorities should want to avoid concessions that might jeopardise them. Pakistan having no such industries to protect and there being no definite plans for the production of complex goods, its authorities' only motive for not lowering tariffs was the possible loss of customs revenue. Against this, they may have been persuaded by the arguments that lower tariffs would bring in more imports and more revenue, or that income and, therefore, revenues would increase with the better allocation of resources, or that it was better to raise revenue from other sources, direct taxes especially. The second reason was that India was better able to bargain than most low wage countries. It did not need balance of payments financing and its authorities did not change their industrial plans because the organisations providing aid wanted them to. They had also created the capacity for trade negotiations and could send enough officials with suitable expertise to negotiations, especially the multilateral rounds. In contrast, Pakistan, like many other low wage countries, had few or no specialists in trade negotiations, but relied on its diplomatic staff.

Industrialisation: the alternatives

In theory there are five different ways of industrialising that Pakistan can try or could have tried. One was that of India, partial autarky, to allow the production of complex goods using generally available knowledge and such knowledge as high wage country firms would impart. The second was that of Taiwan and Korea, obtaining the proprietary knowledge from the firms of a high wage country to allow it to produce components of complex goods. The third would have been to imitate Indonesia, Mexico, Malaysia, Philippines and Thailand in attracting investment by high wage country firms without acquiring proprietary knowledge to produce simple goods, to carry out simple steps in the production of complex goods or even to carry out some complex steps. In the fourth Pakistani firms could produce components of complex goods, as did Taiwan and Korea at the start, but without obtaining proprietary knowledge. Finally, Pakistan could go on producing simple, low wage goods.

Of these the first was prevented by the lack of capacity for training scientists and engineers needed and is now ruled out, especially for capital

goods production, by the low levels of protection allowed by international agreements. Even if the education system had been adequate, the second way of industrialising would have been ruled out before the 1980s because Pakistan, like India, did not have the equivalent of a Japan, whose firms expanded their own production by transmitting proprietary knowledge to the firms and joint ventures of Taiwan and Korea to let them produce more and more complex goods and also become original equipment manufacturers. As the firms of high wage countries have transferred more production abroad, they have transferred also some of their proprietary knowledge to firms in countries with the requisite education systems, like India and, most of all, China, primarily for access to their domestic markets, but also to produce exports. But firms reluctant to transfer proprietary knowledge usually choose the third way, investing directly in low wage countries.

For Pakistan the third way amounted to competing with the countries mentioned for the foreign investment of Japan, the US and Europe, and there it had several disadvantages. One was that Pakistan was not as accessible, especially for shipping. Proximity to the oil producing countries of the Middle East was no advantage in the production of components of goods made by Japanese or US firms. Another is that the countries of East Asia are pleasanter for people from high wage countries to live in. Pakistan is a mostly arid country with a harsh climate, no beaches to compare with those countries' and conservative mores. As long as there were large numbers of unemployed low wage workers in these countries, notably Indonesia, Mexico, Philippines and Thailand, Pakistan could not expect to attract much FDI from high wage countries.

The fourth alternative, producing components of complex goods for foreign firms without acquiring proprietary knowledge, is still possible. Since only generally available knowledge is involved, the activities remain low wage activities and the technically trained workers' wage rates remain below those of their counterparts in the high wage countries. But it allows more diverse exports and, perhaps, the local production of inputs for them, and may with time lead to the beginnings of the second alternative, making components that are technically more complex and demanding of more highly trained workers, which should increase incomes to the extent that these workers are available. Workers in these activities would be paid more than untrained workers and their wages would rise as the extent of these activities increases. Eventually some firms may begin to acquire more advanced proprietary knowledge through partnerships and joint ventures. Parts of economies like Thailand's and Vietnam's consist of such activities. There has been a beginning in Pakistan with firms providing information technology and data processing services using generally available knowledge, though such activities will not progress enough toward the

second alternative to have a noticeable effect on income per head without improvement of education in the country.

What remained was to increase income and employment by increasing the output and variety of simple, low wage goods that are not components of complex goods. Pakistan's industrialisation has followed this pattern since the collapse of the Korean Boom in 1953. Value added has been low in relation to investment and wages have had to be kept from rising relative to those of competing low wage countries, which has been done by repeated devaluation.

The Economic Consequences of Liberalising Trade

More could not be expected. The little importance that education had been given had kept the level in Pakistan among the lowest in the world, which it still is. Expenditure on education by the state was only about 1.5 per cent of GDP, with the consequence that over the thirty years from 1951 to 1981 the literacy rate only rose from 16 per cent to about 26 per cent. University graduates in science and engineering were at about entrance level of the better high wage country universities, post-graduate training or research in these subjects was negligible and there was practically no vocational training for industrial workers. Families who could afford it sent their children to private schools, many of which were run by foreigners, often Christian missionaries, or they sent them to schools in Britain and, if possible, to universities in Britain or the US. Few who graduated in science or engineering from foreign universities returned.

If Pakistan was to follow any of the three patterns of industrialisation just described in the next few decades, its political leaders and economic advisers would have had to realise that postponing the wider spread of education at higher standards to after an industrial "take-off" was to postpone both forever. But they also had to have more government revenue, for an improvement on the scale needed would only have been possible with an expansion of public education and a corresponding increase of government expenditure, and one constraint on that was the exceptionally low ratio of government revenue to GDP when compared to other countries. During the 1970s, consolidated federal and provincial government revenue, including non-tax revenue, was 16–17 per cent of GDP, around a third of the proportions of high wage countries. With so little revenue any increase of spending on education by the state would have had to be met by reducing other important expenditures, and for that it would have had to be accorded greater importance.

Private firms, though free to provide education at all levels, do not substitute for a state system in Pakistan, for they only meet effective demand, the demand of families that have the money to pay for it and live where their children have access to it. There is no lack of desire for education; almost

all families want their children to go to primary and secondary school. Boys have priority, but, outside some tribal areas, families normally want their girls to go to school too. But most desire is not effective demand; most families cannot pay for schooling or they live where none is available. In some rural areas and poor parts of towns such schools as are available are run by charitable organisations, some secular and some religious. Higher education is also provided by the state at low cost, but, apart from a few new elite institutions, the amount and the quality are still low in comparison to other countries. Some of the demand for higher education in sciences and engineering is met by study abroad, especially in the US and Europe, which is how most of those whose education is of Western standards have obtained it. But this is mainly confined to upper income families, for scholarships and grants for study abroad are rare, and those who can afford it for their sons and daughters create little domestic demand. In contrast, India's authorities made sure of the country's ability to provide a high standard of education in sciences and engineering with its institutes of technology, despite the many economists who as late as the 1980s regarded it as misplaced investment.

Families attach enough importance to education for the inadequacy of the public system to have caused private education to grow to account for larger proportions of education at all levels than in most comparable countries. In 2005 the proportion of children of ages 5 to 9 in school was 68 per cent, the same as Nigeria and well below the levels of countries like India, with 89 per cent, and Egypt, Philippines and Sri Lanka, all with 94 per cent or more.²²¹ Pakistan's proportion went up to 77 per cent in 2017, despite a slight fall in the number of public primary schools after 2012, mainly because private schools increased their share of pupils from 30 per cent in 2005 to 39 per cent in 2017.²²² For comparison, the proportions of children in private primary schools were 2 per cent in Sri Lanka, 7 per cent in Egypt, 8 per cent in the Philippines and 17 per cent in India.²²³ For the age groups 10 to 16, the Middle, High and Higher Secondary schools, 33 per cent of the enrolment was in private schools, which constitute the majority of schools and have the majority of teachers at each level. Firms providing private education in Pakistan are among the biggest employers in South Asia.

Of the 185 institutions counted as universities with 1.46 million students in 2016-17, 75 were private with just under a fifth of the enrolment. The education authorities readily recognise higher level degrees that meet their standards and institutions offering recognised degrees have proliferated.

²²¹ Lynd, "The Education System in Pakistan," Chart 6.

²²² Government of Pakistan, Ministry of Federal Education and Professional Training, United Nations Education, Scientific and Cultural Organization, and United Nations Children's Fund, *Pakistan Education Statistics 2017*, 8,13, Table 1.2.

²²³ Lynd, "The Education System in Pakistan," Chart 12.

Below university level they include technical and vocational training institutions with 365,000 pupils and “degree colleges” with 965,000, mostly public. Apart from them are businesses for learning a variety of subjects, information technology being especially popular. Altogether it is not enough for training the engineers, scientists and intermediate level workers needed to change the pattern of industrialisation and, overall, the private provision of education has not raised the international ranking of Pakistan’s education, though it may have prevented it from falling.

Political leaders and government officials seem more aware now of the need for more education than they were then, in great measure because the popular demand for it is evident. But the lack of revenue constrains them at least as much as before. Through the late 1980s and early 1990s the share in GDP of consolidated federal and provincial government revenue had been 17-19 per cent, but it then fell to around 13 per cent before rising to about 15 per cent after 2014. Consolidated government expenditure has mostly been below 20 per cent. For comparison, general government tax revenue in India has been 16 to 17 per cent of GDP and expenditure around 24 per cent.²²⁴

Because government revenue is insufficient budget deficits have been a large part of the country’s excess of expenditure over production. They were less than 4 per cent of GDP in only three years since 1980 and have mostly been over 6 per cent. Pakistan has an exceptional need in its irrigation system, the biggest in the world, on which much the greater part of its agriculture depends, and the maintenance alone of which is costly. Foreign debt service is another item that must be met. In 1974-5 it was Rs. 2.3 billion (14 per cent of government expenditure), as compared to Rs. 55 million for education, in 1982-3 it was Rs. 10.6 billion (18 per cent of government expenditure), as compared to Rs. 780 million on education, in 2005-6 it was Rs. 104 billion (7 per cent of government expenditure), as compared to Rs. 46 billion for education, and in 2017-18 it was Rs. 1,950 billion (26 per cent of government expenditure), as compared to Rs. 829 billion for education. Then there are the military outlays, which are large in relation to the budget, though not necessarily in relation to GDP, and which are a function of geography and history. Other functions of the government have been impaired by the chronic shortage of revenue, in particular the construction and maintenance of infrastructure, provision and regulation of electric power, provision of security, the application of the law through the courts and protection of public health through water treatment, waste disposal and disease control.

Education is one of several examples showing how trade liberalisation, rather than make Pakistan’s economy more competitive, has done the opposite by preventing government revenue from increasing relative to GDP. One

²²⁴ International Monetary Fund, “World Economic Outlook Database April 2011.”

reason for the fall in the share of government revenue in Pakistan was that import duties were lowered before alternatives could be put in place. In the mid-1950s and as late as the early 1990s taxes on international trade had been, as was typical of low wage countries then, the biggest single source of government revenue; they provided over 40 per cent of tax revenue, as compared to direct taxes, which provided around 15 per cent. From the 1980s on their part of tax revenue declined to below 10 per cent in 2012-13, when tax revenue, itself, was down to 10 per cent of GDP. Since then their part has risen to about 14 per cent. There was also a fall in the share of non-tax revenue for a variety of reasons. In addition, various governments have reduced or waived several taxes and caused a loss of revenue that Pasha and Ghaus-Pasha calculate was in 2010/11 equivalent to 34 per cent of total revenue, or about 3 per cent of GDP.²²⁵ If that revenue had been collected Pakistan's ratio of government revenue to GDP, though still low in comparison to other countries, would have risen above what it had been for decades.

In theory, the loss of import duties should have been compensated by increasing income taxes and applying a value added tax. New and higher taxes on the scale needed, aside from being politically difficult, as in any country, require budgetary outlays on improving the relevant administrative capabilities. First should come the expenditure on the tax administration and then, when the additional revenues have been assured, comes the lowering of import duties. There was little administrative capacity needed to capture incomes, other than declared salaries from firms and government service, and there was the political problem that wealthy landowners had the political power to keep agricultural income exempt. About 1.5 per cent of the population files income tax declarations. Outside agriculture, income tax is mostly collected either as withholding tax on items like salaries, yields on financial assets and export proceeds, or is calculated on income imputed from indicators like electricity and telephone bills.²²⁶ Taxes on profit were, in theory, not exempt, but there were many tax incentive schemes or holidays and the tax authorities' capacity to investigate evasion was small. Revenue from direct taxes has risen faster than GDP only since 2014. A General Services Tax has been introduced as a precursor to a value added tax and should replace excise, which is the main indirect tax on domestic sales but is confined to specific goods. But it requires producers, importers, wholesalers and retailers to declare their sales and retailers, in particular, have resisted, which has also hindered the reduction of the evasion of taxes on business profits.²²⁷ Yet, the composition of tax revenue has changed since the 1980s; direct taxes have risen from around 18 per cent of tax revenue in

²²⁵ Pasha and Ghaus-Pasha, "The Future Path of Tax Reforms in Pakistan," Table 7.7.

²²⁶ Pasha and Ghaus-Pasha, 197.

²²⁷ Ahmad, "Why Is It so Difficult to Implement a GST in Pakistan?"

1990/91 to around 35 per cent in 2017-18 and sales tax from 18 per cent in 1998-99 to 35-39 per cent after 2011.

Those working for international institutions and giving economic advice who demanded that import duties be lowered and who realised also that it would result in a loss of government revenue did not think the loss a reason for delay; they did not regard the effects on the country's public finances as their responsibility, especially not if there were political considerations. In terms of domestic politics it is usually easiest to increase import duties. But doctrine was and still is that import duties distort trade and it is up to the country's authorities to find ways of raising revenue that do not do that. It follows from the overly simple notions discussed earlier of world prices and allocation of resources and does not take account of the economic functions of government expenditure of the sorts described. It would be sounder to argue that shortages of government revenue give rise to inefficiency and that suitably chosen import duties improve efficiency by providing revenue. Such duties do not cause any inconvenience to the high wage countries if they do not discriminate among countries but conform to the most favoured nation principle, although the principle is routinely flouted by the high wage countries, themselves, for instance by imposing import duties on Pakistani textiles while textiles from countries, like Israel and Turkey, with higher incomes per head, enter free of duty.

Two more examples, transport and electric power, illustrate how costs to the economy would be reduced if the government increased its spending to improve the country's infrastructure. Most goods are transported by road and, because the highways are congested and often in bad repair, it is costlier and slower than need be. The road system roughly tripled from 1980 to 2012/13, in addition to which much was upgraded, but the numbers of buses and trucks registered increased by a factor of four and passenger cars almost tenfold. Some of the increase of road traffic could have been avoided if the railway system had not shrunk so much; its freight, measured in tonne kilometres, decreased to roughly a twentieth. Electric power supplies have not kept up with demand, though they increased by a factor of six. Power cuts for industrial and commercial producers, as well as households, have become so frequent that those who can afford it have their own generators despite the cost. To attract private firms to invest in power production the authorities set the price of electricity relatively high and adjust it according to some international norms. What was meant to improve the finances of the power companies often has the reverse effect; it increases theft through illegal connections because electricity is a necessity and it induce various parts of the government to delay payment for their electricity consumption because of budgetary shortages. In both cases, improvement requires more expenditure by the government, though other changes are needed as well,

one being revoking a covenant with the World Bank forbidding the state, but not the private sector, from investing in thermal power generation.²²⁸

Besides being a direct cause of inefficiency, lack of government revenue results in the deterioration of the administrative system, itself. Curtailing the salaries of government employees is an easy way to keep government expenditure down or, alternatively, to increase staff and non-salary expenditure on infrastructure, schools, subsidised healthcare and so on. But, for efficiency, the administrative system must have suitably competent staff. Curtailing salaries, if repeated too often, determines pay by revenue, rather than by qualifications and responsibilities, and results in a deterioration of the qualifications and abilities of the staff. As economies grow and technical advances take place the functions of the authorities become more complex and the administrative system needs abler people, often with more specific expertise. Experts can often be hired for specific tasks from outside the administrative system, but that does not necessarily reduce the expenditure. Hired experts are likely to be paid more than civil servants, often several times as much. Besides, it is the civil servants who must determine what the tasks are, evaluate their advice, especially when, as often happens, the experts differ among themselves, and carry out the recommendations or adapt them to political considerations. In addition, most of the work of the administrative system is routine, requiring competence and familiarity with the matter, not specific tasks that require expertise from outside.

Poor prospects of improving pay through effort, qualifications and seniority are a stimulus to corruption, which is now common in Pakistan. It was not common at the start; before the 1960s the extent of the corruption was probably no greater than in the US. In particular, politicians, though they may have shown few scruples in their political dealings, did not enrich themselves. But when corrupt leaders did take power, it spread from the top. At the time corruption was taken lightly by many economists and political scientists, including foreign advisers, who even approved of it as the workings of market forces circumventing undesirable administrative regulations. In the country many could see further than that, especially the consequences for the social system and economy as a whole, which are pervasive and too evident to need describing here.

Corruption has taken on such proportions in most low wage countries that it is now often used as the main explanation of the failures of economic development. As an explanation it seems easy to understand, whilst the reduction of corruption seems slow and difficult. It also makes the source of the problems the country, itself, not the economic advice its authorities receive, the measures they may have been made to take or the trade policies

²²⁸ Chaudhry, "Pakistan Indus Basin Water Strategy," 259.

of other countries, and it makes for a vicious circle, economic failure provoking corruption and corruption hindering the functioning of the social system, including the economy.

Judging by the example of much of Europe in the eighteenth and nineteenth centuries, escape from this vicious circle has to begin with economic improvement. A common belief is that authoritarian governments can suppress corruption, reform institutions and pass desirable laws. Usually, though, such governments rely on the support of interest groups, even if they do not represent them, and are, therefore, liable to be corrupt themselves. But even the most honest must fail if it cannot raise incomes and employment fast enough and it must do this without failing in the country's external debt obligations or breaching its various agreements on trade liberalisation, foreign investments, capital flows, etc.. In one respect the suppression of corruption is more difficult than it was in Europe because much of the wealth acquired by corrupt means is transferred to accounts in high wage countries and it is rare that the authorities of these countries question these inflows or help the authorities of the countries from where they come to find and retrieve them. Bank secrecy concealed much of this money and nowadays the US is under no obligation to report accounts held by foreigners.

It can be asserted that Pakistan's authorities should have taken measures to increase tax revenue before they were obliged to lower import duties; they should have known they could not escape complying with the demands of the high wage countries and international organisations, such as the GATT/WTO, the World Bank and the IMF, of which they knew the doctrine regarding trade barriers. Had they done so, they would have had more revenue when the duties were lowered and more could have been done to increase the efficacy of the administrative system or prevent its decline and to reduce corruption. Revenue now, too, would be greater than it is, both because the tax administration would have more and better trained staff and because the process of improvement would have been more advanced. But, if Pakistan's authorities should have known in advance that they would eventually be obliged to lower import duties, the high wage countries that demanded trade liberalisation should have shown the same foresight when giving their economic advice. It would have been foresight of some decades, for that was the time that would have been needed to create new and modern administrative and tax systems. As a typical low wage country that had been ruled by another country, Pakistan had begun with an administrative system limited to what the ruling country's authorities had judged essential and for which they provided the highest level of staff. It was not an administrative system designed for economic development and social change.

The Exchange Rate and its Consequences

Attitudes towards the exchange rate have changed since the time of the Bretton Woods system. It used to be that devaluation was regarded as a sign of failure or inferiority, and that as much in Pakistan or Madagascar as in the UK. In some countries this feeling continued into the 1980s. Economic doctrine has, nevertheless, always asserted that persistent balance of payments problems can only be solved by devaluation and, with time, the doctrine that the market should determine the exchange rate became the orthodoxy. As upholder of the doctrine, the IMF, which had been created to prevent exchange rates changes if possible, became a proponent of floating exchange rates or of, at least, prompt devaluation. Officials of central banks and other monetary authorities in low wage countries who learnt the doctrine through the IMF's courses or negotiations with its staff, could demonstrate that they no longer had out-dated attitudes by their willingness to let their currencies depreciate. Most low wage countries that devalued once had to devalue repeatedly and their officials and economic advisers, who may have yielded reluctantly at first, gradually lost their inhibitions. Rejection of their earlier attitudes came to be a sign of maturity. They then had to believe that nominal devaluation was necessary to offset domestic inflation and even that devaluing so much that the currency was "undervalued" made the economy more competitive.

As long as the authorities of a low wage country that produces no complex goods wish to be considered orthodox, repeated devaluation is an economic certainty. That Pakistan should ever revalue its currency was out of the question. This was demonstrated in the 1970s, when inflation was high in the high wage countries and Pakistan's authorities complained about having to "import inflation". By the logic of the argument for devaluation, the remedy would have been to revalue, but that a low wage country like Pakistan should do so was never considered. Yet reasons to devalue were always found and the inflationary effects were ignored or denied.

Pakistan's first devaluation occurred in 1956, when low prices for its principal exports, cotton and jute, and high demand for investment goods put the country in balance of payments difficulties. From then until 1972 the official exchange rate of the rupee was Rs. 4.76 to the US dollar. While the EBS lasted the exchange rate for imports varied between the official rate for those deemed essential to about 250 per cent of that rate for imports paid for entirely with foreign exchange bought with bonus vouchers. Exporters of manufactures received about 150 per cent of the official exchange rate and other exporters usually the official rate. Abolition of the EBS required that the official exchange rate be changed and the new rate of Rs. 9.61 to the dollar was kept for ten years. From 1982 on the rupee has been depreciating

more or less continually to reach about Rs 100 to the dollar in 2014 and about Rs 160 to the dollar in mid-2019, a rate of 7.6 per cent p.a.. The GDP deflator rose at about the same rate, 0.46 per cent p.a. faster, but slower than worldwide inflation.

Wages

Data on how much unskilled workers and skilled manual workers were paid from Partition to now are not good. An index of wage rates used to be compiled in the 1960s, but since then the only more or less regular data on overall wage rates have been provided by the International Labor Office (ILO), which publishes figures for mean monthly wages from 1969 on with a few gaps. The Pakistan Bureau of statistics publishes figures for the average daily wages in October or November each year of construction workers in various cities, though they were only available from 1975 to 2007. Overall, the two series are roughly in conformity over time, though some vagaries, especially of the latter series, make the figures for some years seem unreliable. It is mainly the ILO figures that are used here. Assuming that their trends do indicate changes in the pay that unskilled workers receive in general, three comparisons can be made: with foreign currencies, in this case the US dollar, with prices and with income per head.

When industrialisation started in 1954 real wages, as given by the official figures, rose at first. Factory work was new and workers available for it were scarce. Part of the reason for the scarcity was that the demand was restricted to where the investments in cotton manufacturing were being made, mainly Karachi and some areas of the Punjab, and the urban populations were small. Probably more important was that the new textile industry was displacing the handweavers, who had been one of the main producers of cloth in the country and did not willingly give their vocations up. They were not being displaced by price competition; comparison of their output of simpler varieties of cotton cloth with the prices of the cloth produced by the new power looms shows that they could have competed, even lived well during the years when cloth prices were especially high. Their disadvantage was that hand spinning could not compete with machinery and the industrially produced yarn was kept for industrial weaving. They were also at a disadvantage in that they needed to act as cooperatives of industrial groups to obtain the dyes, chemicals, designs and marketing arrangements to sell their products. When they were so organised they prospered. If they had been organised and if they could have obtained their yarn on competitive terms, the handweavers could have competed with the new industry on equal terms and the country would have had its basic textiles more cheaply without spending a lot of foreign exchange on displacing its existing production capacity. Instead of a complete ban on textile imports, suitably

high tariffs would have allowed imports of varieties of cloth that were lacking in the country whilst protecting the handloom industry. As it was, the higher wage in industry did not last.²²⁹

Real wages, as measured by the ILO's figures deflated by the GDP deflator, had no durable increase until they rose in the years before 2017 to reach their highest level ever. Before that they only fluctuated; they were the same in 2001 as in 1975, which is also what the series for the daily wages of construction workers shows. Both series agree that real wages were a bit higher in the years 1977-1979, again in 1987-1990 and 1995-1998. Despite gaps in the ILO series from 2003-2006 and 2010-2012, it seems that real wages rose consistently for several years before 2017. They may have risen in 2018 as well, but a figure for the GDP deflator for that year was not ready.

That real wages were prevented from rising by the depreciation of the currency can be seen from the ILO figures for the monthly nominal wage converted into constant dollars, which shows that it did not rise above earlier levels before 2017. Converting the nominal wage into dollars at the average exchange rate for the year and deflating by the US GDP deflator shows that, until then, it fluctuated below its levels of 1978-80. As nominal wages rose domestic inflation prevented prolonged increases of the real wage and exchange rate changes stimulated domestic inflation. This is being repeated with bigger fluctuations by the devaluation from Rs. 110 to the dollar in mid-2018 to 160 in mid-2019; from their highest levels the constant dollar wage and the real wage have been lowered by an exceptionally big devaluation, which will be followed by a correspondingly high rate of inflation.

Reducing the nominal wage relative to foreign prices and wages in other countries is a standard theoretical reason for devaluation and is necessary if high wage rates make the economy insufficiently competitive. But in Pakistan wage rates have never been the explicit motive. If they had been there would presumably have been better data on wages. Devaluation seemed necessary, not because of high wages, but because of the belief that trade deficits demand it. There also seems to be no grounds for taking it as far as it has been in Pakistan. The advocates of devaluation never made the relevant comparisons; no attempts were made to assess if exports would be increased by lowering wage rates relative to foreign prices, or to what extent reducing the consumption of those whose income would be lowered by devaluation would improve the trade balance or whether or not curtailing demand would not be more effective. If there is production capacity that was not profitable before and becomes profitable because of the devaluation, it is not competitive despite Pakistan's low wages and that may be because the firms are inefficient or there are obstacles, like power shortages and bad

²²⁹ Khan, "What Has Been Happening to Real Wages in Pakistan?"

infrastructure. In such cases devaluation provides temporary relief at a high economic and social cost, but does not remove the inefficiency.

One indication that the wage rates have not been the cause for any lack of competitiveness is the greater unevenness of income distribution that results from devaluation. If wages fall relative to income per head, there is a presumption that profit margins have increased and so have incomes in activities other than production of exports and import competing goods. If wages have risen too high for competitiveness, why have other incomes not been lowered? For there to be an argument for devaluation wages should for some time have been higher relative to income per head when compared to periods when they seemed competitive. From 1969 to 2000 the ratio of the ILO wage to income per head fell from being mostly above 2.0 to about 1.6. From 2000 on the national accounts were put on a new base and prevent direct comparison of the ratio before and after 2000. But from 2000 to 2017 the ratio of the real wage to income per head was roughly constant, with a small rise for some of the years in between. Much the same pattern for the years 1975 to 2000 occurs with construction worker wages; they fell relative to income per head. But the ratio also fell from 2000 to 2007, the last year for which these figures were available.

Inflation

Pakistan's authorities seem not to have stated explicitly their reasons for no longer trying to keep the exchange rate fixed, but at the beginning one practical reason may have been that trade barriers were being reduced and that they thought devaluation would help industries that had needed protection. Otherwise the reasoning seems at first to have been that regular depreciation of the currency was what low wage countries did to improve the balance of payments. After that, if domestic inflation exceeded the chosen measure of price increases outside the country, the currency had to be devalued to offset the difference. The comparison was of changes of indices, not of actual domestic and foreign prices.

Another reason for letting the currency depreciate was that it seemed to be an adjustment to the market price. It began to operate in 1985, when the authorities allowed foreign exchange from remittances by Pakistanis abroad to be sold on the market and taken out of the country as foreign exchange bearer certificates (FEBC). As the only legal way to transfer money abroad, apart from the transactions allowed under the current account, foreign exchange bought in this way was bound to cost a premium over the official rate determined, in the first instance, by overall demand and the quantity of money. Other influences on the premium were expectations about the economy and the exchange rate and, since the market was exempt from controls, the desire to get "black" money, i.e. money acquired illegally or

concealed to evade taxes, out of the country. On the argument that the exchange rate was a price and should, like any other price, be determined by the market, the existence of a premium was interpreted as a sign that currency was overvalued, even though the main immediate reason for it, the money supply, was controlled by the authorities. A separate price for foreign exchange bought with FEBCs was, however, a multiple exchange rate practice contrary to the IMF's rules and had eventually to be given up.

As to be expected, continual devaluation caused inflation; from 1982 to 2000 the increase of domestic prices as measured by the GDP deflator was over 400 per cent, roughly the same as the price of the US dollar in terms of rupees. But the central bank, the State Bank of Pakistan, conformed to the standard doctrine of central banks that the cause of inflation was the money supply and devaluation the necessary consequence. Some connection between money supply and inflation is to be expected, for there are two mechanisms that give one. In the one, an increase of money supply causes total expenditure to increase. Then, if the exchange rate is set by the authorities, they react to the worsening of the trade balance by devaluing, and, if it is set by supply of and demand for foreign exchange, the greater demand raises the price. Domestic prices then rise. In the second, devaluation or a rise in foreign prices in terms of the domestic currency induces the authorities to increase the money supply rather than provoke a contraction of domestic activity. In either case the direct cause of the domestic price increases is, not the money supply, but the relative increase of foreign prices.

If the authorities are persuaded that the direct cause of inflation is the quantity of some form of money, a self-perpetuating cycle of devaluation, inflation and increases of money to accommodate the inflation is almost assured. If, as was the case in Pakistan for a while, they set the exchange rate from time to time, the "crawling peg", they devalue to offset the amount by which domestic inflation exceeds the chosen measure of price increases outside the country, the comparison again being, not of prices, but of changes of indices. Since curtailing the money supply does not stop the ensuing inflation, the monetary authorities accommodate the higher prices with more money, the second mechanism mentioned above.

If, as has been Pakistan's case for some time, the exchange rate is mainly determined by the market, the monetary authorities' first concern should be regulating the money supply or credit according to rate of inflation they wish to try for. They can try for no inflation by reducing the relevant quantity to eliminate demand in excess of the supply of goods and then changing it in proportion to the supply, with due adjustment for changes of foreign prices, though restricting credit that much is thought likely to deter investment or

even cause a fall of output. In some of the major economies the central banks prefer to keep inflation close to but below two per cent per year, which they believe experience has shown not to be too restrictive. Inflation in Pakistan has averaged about... per cent a year, so such rates are regarded as unrealistic for the time being. Pakistan's central bank, the State Bank, has chosen to keep inflation at five per cent per year. Inflation has rarely been that low in Pakistan, but five per cent a year is more than most usual measures of the rate of increase of world prices. What follows, is that the increase of the money supply increases demand for foreign exchange and raises its rupee price as described in Chapter 3, unless there was an unanticipated increase in its supply. Depreciation is followed by inflation, though not necessarily at the rate intended. Here, too, the second mechanism operates, though the quantity of money is set by an inflation target.

Empirical evidence for the effect of the exchange rate on inflation is provided by the data of the econometric study carried out in 1996 for the State Bank, though that was not the study's purpose. What the study purported to conclude and was intended to show was that the 'money supply would appear to be a key determinant in an economy'.²³⁰ But, in its Section 1, the description of events over the period of 1970-95 shows that high inflation in Pakistan is always associated with devaluation or high international inflation and low inflation with a stable exchange rate and low international inflation and that the association with money supply is not as close; in certain periods (1977-79, 1982-83, 1992-93) inflation and money supply did not move together. In the study's diagram in which the consumer price index is plotted and the rupee prices of tradable goods (a proxy for international prices) over time (Fig.1.b), the CPI follows the prices of tradables with a short lag. No such relation is apparent in the diagram of the CPI plotted against the money supply; rather it is the contrary movements that stand out.²³¹

Other statistical studies provide evidence that the exchange rate causes inflation, of which only two purporting to show that depreciation is not a cause need be discussed. One, from 2002, by Choudhri and Khan asserts that concerns that devaluation may cause inflation in Pakistan are 'misplaced'. If that is to say that depreciation does not cause inflation, the result is illogical and must be explained. One explanation is that both an index of foreign prices and the exchange rate are among the independent variables of the study, though the latter works through the former and is, therefore, redundant. Another explanation is that the study is limited to the short term, a year at most, which makes it doubtful that it can detect causation

²³⁰ Nasim, "Determinants of Inflation in Pakistan," 1.

²³¹ Nasim, fig. 1.d.

with confidence when depreciation and inflation are continual. Later, in 2006, Khan and Schimmelpfennig tried a number of models and concluded that the leading indicators for forecasting inflation, apart from lagged inflation, were the growth of private credit and, perhaps, of broad money. This is consistent with the two exchange rate mechanisms described above, but is not an economic mechanism and none is provided by the study. Instead, as often with such statistical exercises, the study's specification was chosen to give the results wanted. It begins with a list of variables considered plausible, rates of change of prices, of various measures of money, of GDP, of the nominal exchange rate, etc., and proceeds by using advanced techniques to try various combinations of the variables and lags to find which give the results that conform best to what is being sought. One of the combinations that were tried included the exchange rate, which was excluded from the leading indicators because it had the wrong sign. None of the three studies follows the correct procedure for testing the relation between the exchange rate and inflation, which would have been a direct statistical test using some index of domestic prices and indices of import and export prices with suitable lags, and not rates of change. If the money supply is to be explicitly included in the test, it cannot be treated as an independent variable because of the second mechanism described above.

Capital flight

The reasons for devaluing given here are only remotely connected to production or prices of manufactures, if at all, but they have given rise to another self-perpetuating cycle, that of capital flight and depreciation. People and firms convert domestic financial assets into foreign ones to reduce risk by diversifying their assets, especially the risk of losses caused by depreciation of the currency when they think it is likely, to avoid taxes and to conceal black money. Removing some of the restrictions on holding foreign assets and eliminating the premium on FEBCs presumably increase capital outflows, other things being equal, and thus cause the currency to depreciate more.

Earlier, when citizens not resident in another country were forbidden to hold assets abroad only a few people could spread risk by holding some wealth as foreign assets, mainly those who could acquire foreign currencies easily, especially exporters and importers, who could misinvoice, and those who were rich enough to pay a premium for converting rupees on the black market. Those who were not rich were at a disadvantage because they did not travel as much and would have been less able to look after or use any foreign assets. This changed with the rise of employment in the Middle East and the introduction of the FEBC. Not only was it legal to take money out of the country, but more people who were not rich could also expect to be frequently outside Pakistan and to be able to look after their assets as well as

to use or to add to them. In time it became normal for better off middle class families to have foreign bank accounts and to invest in assets issued by firms and governments of high wage countries. They were helped by a decline in the cost of transferring money; while FEBCs were in use the cost included the premium, but the adoption of a unitary exchange rate reduced the cost to the normal transaction costs and obtaining the central bank's permission.

4. POVERTY AND PESSIMISM

The distribution of income

Despite economic growth and higher incomes per head, the poor in low wage countries mostly remain poor and they know they have little prospect of a future of rising incomes. They know it from long experience; they have not seen the changes of pace of growth of employment and expansion of education that are the preconditions of a general betterment for them and their children. They share employment and unemployment among themselves, and the attendant insecurity. What they see as the inability of the economy to provide the education they want is to them also the reason there is virtually no social mobility for the families of untrained workers; the small chance of the children of a peasant, street vendor or mechanic have of becoming doctors, engineers, scientists or lawyers.

Much the same can be said, with suitable adjustments, for the greater part of the population. In a low wage country like Pakistan changes in the nominal incomes of the different segments of the bulk of the population do not vary much over the medium term. Fixed incomes, wages and salaries, move more or less together, with the possible exceptions of the pay of skilled people who can emigrate easily, managers of firms, high ranking government officials and politicians. Incomes that are not fixed, such as those of small shopkeepers, vendors, craftsmen and occasional labourers, are not likely to change very differently. This is not true of the incomes of owners of big businesses or large tracts of land, nor, perhaps, of the more successful in the liberal professions, like doctors, lawyers, architects, etc.. Here the term "nominal wages" is used to represent the incomes of the less wealthy part of the population and can, itself, be represented by an index of wages, if there is a reliable one.

It follows that the less nominal wages in a low wage country rise in terms of the currencies of the high wage countries the less incomes of the bulk of the population can be expected to rise. If the exchange rate is constant and income per head grows, nominal wages may rise faster than prices, which are mainly determined by prices outside the country. It depends on the amount of unemployment and the social and political circumstances of the country, and on it depends the degree of income inequality. If the currency depreciates slowly the state may succeed in preventing a worsening of

income distribution by providing good education and health systems and allowing wages to rise enough. This is rare; about the only case among low wage countries seems to have been Tunisia in the period of around 1970 to 2000. If the currency depreciates rapidly, real wages fall and measures to prevent inflation cause unemployment; income distribution becomes more unequal to the extent that total income does not fall.

Income per head may be rising but cannot be supposed to "trickle down" unless there is a specific mechanism that makes it do so. It can be a combination of fast economic growth from producing goods that are not low wage goods and the spread of the education needed for such growth. This is what happened in economies like South Korea and China, despite the slow rise at first of the wages of unskilled labour. Alternatively, it can be measures by the state to raise real wages or provide support for the poor, measures that cannot work for long unless unemployment is low and tax revenues are adequate. Occasionally, though rarely, both conditions have been met, as in Malaysia, where the poor have shared in the rise of income. Otherwise they are rarely met since virtually all low wage countries have high unemployment. But, in Pakistan, as in many other countries, there was no attempt to make income trickle down; the purpose of repeated devaluation of the currency was to keep nominal wages down.

Poverty programs

When, because of the economic policies or circumstances, nominal wages do not rise in terms of the currencies of the high wage countries the worst of the poverty in a country can only be reduced or alleviated by programs specifically for those purposes. Orthodox economists, especially in the multilateral and bilateral development institutions, have been preoccupied since the early 1970s by the meagreness of improvement in the living standards of the majority of the population in most low wage countries, just as in Pakistan. Poverty programs are, therefore, now among the main activities of these institutions, as well as of a multitude of charities and other non-government organisations (NGOs) and the authorities of the low wage countries have been persuaded to devote more budgetary expenditure to them than they otherwise might have done.

Alleviating poverty is not the same as reducing it. Alleviation improves the living conditions of the poor, e.g. better water and sanitation, preventing pollution of the immediate environment or providing better nutrition, safer housing, etc., without necessarily increasing income, though perhaps lessening obstacles to generating it. Much of it is communal and, though it improves welfare in many ways, the gains are not in the form of greater purchasing power and are only incidentally apparent from the national accounts. Poverty reduction consists of increasing the purchasing power of

the poor, for which income creating activities have to be found or income diverted from the less poor to the poorest.

Apart from often being novel and ingenious, programs to alleviate poverty are too diverse in their objectives and methods to describe here. Many have improved the well-being of the beneficiaries at small cost, and, much of their financing being grants, have added little or nothing to the countries' external debts. Examples are improving primary schools, providing better food to children, protecting ponds and streams of the localities and reducing ailments caused by parasites and mosquitoes, all of which improve the welfare of the beneficiaries over the short and long runs at no cost to others. Sometimes, however, the benefits do have costs to others or are transitory. Microfinancing to open small shops selling necessities may provide some income for the new shop owners, but, unless the quantity of sales increases, it is income diverted from existing shop owners. Improving the yields of small cocoa and coffee growers around the world, which should be poverty reduction, results in lower prices for them without a proportionate increase of sales. In the end the growers may be worse off.

That it is necessary to alleviate poverty creates a dilemma for proponents of orthodox economic theory. If the conditions for free, competitive markets are met, the outcome is, according to the theory, efficient and labour and capital receive their marginal products. Poverty can only be explained by a shortage of capital and the consequent low marginal product of labour. Hence the dilemma; the necessity of alleviating implies there are criteria apart from this kind of efficiency. Wages defined by purchasing power are, then, not satisfactory indicators of welfare, if only because the conditions under which people live, such as the water supply, sanitation and waste disposal, protection against disease, security from crime or the elements, transport and communications are largely beyond the capacity of individuals or even voluntary groups to improve. Such externalities must be dealt with by institutions that have the necessary legal and administrative powers and are accepted by the affected people, municipalities, provinces and the central government among others. Either ethical or political considerations of fairness are unavoidable, or free, competitive markets are best and poverty programs cause resources to be allocated less efficiently and lower total income.

This is unrelated to the possibility, allowed by orthodox theory, that the conditions for free, competitive markets may not be met and that there may then be justification for interference. A condition that may not be met, for instance, is that the marginal product of labour be a sufficient wage for nutrition and living conditions that allow workers to function efficiently. This is a recognised theoretical possibility and is realised in many places. Nevertheless, poverty alleviation to enable workers to work more efficiently is a rarity;

poverty alleviation measures provide nutritional supplements and health care mainly to children, but are not directed to the productivity of workers. Besides, it would be normally be in the interest of the employers to do that, which is to say that, according to theory, the rational employer increases the marginal product of a worker by paying more or providing assistance in kind. Another condition that must be met is competition; low wage countries often have monopolies and oligopolies. Imports can provide competition for producers of tradable goods and the degree of competition can be regulated by tariffs, provided the country has the freedom to do so. Measures may also be needed to have competition among the distributors, who may form cartels. Ensuring competition in the provision of untradables is often not feasible by the nature of the product, for instance electric power, and has to be compensated by regulation. These are things with which proponents of the orthodoxy are familiar and maintain that theory can be adapted to.

These deviations from the ideal conditions for free, competitive markets do not explain why poverty is so persistent. On the contrary, if they did there would be no need for poverty programs; the measures to correct for them are evident from the theory. Unless it is demonstrated that the need for poverty programs arises only when markets are not allowed to function freely, it has to be accepted that, in principle, these programs may be needed even under ideal conditions, for it is the criteria by which the outcomes are judged that are in doubt. Such a demonstration is unlikely since the progressive liberalisation of the western economies over the last four decades has been accompanied by greater income inequality in almost all of them and signs of widespread poverty in parts of the populations where there had been little before, especially in the US.

Casting doubt on the criteria by which the outcomes of free markets are judged casts doubt on the benefits of trade liberalisation. Against this, proponents of trade liberalisation argue that the alternatives cause greater inefficiency. They often point to how badly some countries' economies seem to have been run in the past. But that is not enough. Put so vaguely some argument in support is needed; if the assertion is that no plausible combination of state activity and protection against foreign competition is efficient enough by some acceptable criterion, the histories of the US, most, perhaps all, of Europe and of East Asia contradict it. Otherwise, the argument has to be made case by case; each program of liberalisation ought to be accompanied by a description of the types of inefficiency and their effects and then by an explanation of why the staffs of government and the firms it controls cannot, with the right training and administrative powers, be expected to do better. Proponents of trade liberalisation do not go into so much detail.

Three consequences of the persistence of poverty are discussed here that have not been confined to Pakistan. One is that both the poor and the upper income groups try to emigrate and the latter try to transfer some of their wealth to high wage countries. The second has been resentment, in particular against the high wage countries of the West. The third has been a spread of beliefs in Utopian societies, usually according to some religious doctrine or a return to some idealised or mythical past.

Looking outside the country

The first is common to all low wage countries that do not produce complex goods. The poor with jobs or plots of land know that work is better paid in the high wage countries and the oil exporting Arab countries and those who have neither and cannot be supported by their families have still more grounds for emigrating. Hence the numbers of poor people from low wage countries trying to enter high wage countries illegally, often at the risk of their lives and at great expense, whilst the high wage countries set up expensive, elaborate systems to keep them out. When children of families who have the means get the education to be scientists, engineers, doctors and the like they often get it in the high wage countries and often stay there. Their families consider it prudent to keep some of their wealth abroad as well.

Castley points out how differently the northern countries of West Europe and America, on the one hand, and Japan, on the other, have behaved towards this kind of immigration. The former at first allowed immigration of workers from low wage countries and protected their manufacture of simple goods, whereas the latter did not allow such immigration and transferred production to other countries.²³²

The former allowed immigration during the first three decades after the Second World War because their economies grew fast enough to have labour shortages. Rather than reduce the production of simple goods, textiles in particular, and let wages rise faster, they protected this production and kept wage growth slower. At first the immigrants into the European countries came mostly from the poorer southern European countries, Greece, Italy, Portugal and Spain, and, in the cases of the imperial countries, from their empires. By the mid-1960s immigration from Egypt, Turkey and East Asia became common as well. In the US they came from Latin America and from Europe. During this time the production of simple goods contracted and wages in southern Europe rose. In the high wage countries the need for untrained immigrant workers declined. At the same time, in the low wage countries, the failure to create enough employment and to raise incomes, along with the greater awareness that came with emigration of the possibilities in the high wage countries, increased the numbers of people trying to emigrate from these

²³² Castley, *Korea's Economic Miracle*, 303.

countries. That immigration has been and is used to keep wages in the high wage countries from rising faster can be seen from the common argument that workers whose origins are in these countries would not do the menial work of cleaning streets, removing trash and the like, much of which is now done by immigrants and their descendants. This work used to be done by native workers and some still is, so more would presumably be done by them again if the pay were better and the investment were made to make the work less unpleasant. In Japan and Singapore, both known for their cleanliness, the work of this sort has been done by people from there.

Without immigrant workers the individual Japanese firm could only increase its output by investing in equipment that raised output per worker, by investing abroad or by concentrating on production that required its more advanced knowledge and transferring the rest abroad. In practice, Japanese firms did all three. As already described, they at first transferred the production of simple goods and then various parts of the production of complex goods at first to firms in Taiwan and, more especially, Korea and, later on, to other countries, like Malaysia and Thailand, often setting up subsidiaries or joint ventures. Until the mid-1980s wages in Japan rose quickly and steadily, which is what was to be expected when producers invested in equipment for making more advanced goods that allowed a greater value of output per head.

According to how much they kept producing simple goods domestically when they could have imported them from low wage countries, the high wage countries of Europe and America forewent any income increase from that source. Their authorities were more concerned with the problems that closing down industries could cause. Since the high wage countries acted in concert, their protective barriers against the simple manufactures of the low wage countries reduced the prices paid to the producers received below what their own producers got or would have got. As their own manufacture of simple goods gave way to imports, they gained income in the form of tariff revenue or lower prices for importers.

Once the process of transferring production to the low wage countries to reduce wage costs has become familiar to high wage country firms, its speed is only limited by the firms' investment plans and, if the transfer is of complex production, by the numbers of suitably trained workers in the low wage countries, unless the authorities of the high wage countries deter such transfers by protecting domestic production against imports or the low wage countries have political problems or their authorities restrict investment. Protection against the imports of simple goods by the high wage countries was intended to deter or delay the transfer of production and did so. In contrast to the firms of the western high wage countries, Japanese firms hastened the transfer of textile production to Korea by investing there,

themselves, and importing the output. Japanese firms were also quicker to transfer more complex components or steps of complex goods production after transferring the production of simple goods.

The transfer of production by Japanese firms to their nearby countries had no counterpart in Europe and the US, whose neighbouring low wage countries still produce simple goods and little else. From the 1970s to the 1990s, while production in the East Asian economies for the high wage countries was growing rapidly, such production as was transferred from Europe and the US to their neighbours in Africa, Asia and Latin America was almost all confined to simple goods and smaller in total value. For instance, Mexico's so-called *maquiladoras* produce a variety of simple goods, often with too little investment in fixed equipment to deter them from moving to other countries. Tunisia's so-called "off-shore" industries are mainly garment assembly operations, the part of making clothing that still needs to be done by hand. Such operations exist in several low wage countries, ranging from Bangladesh to Honduras, for it employs untrained workers and needs little investment. Mexico does have a motor car industry, mainly plants of big European and US firms, and Tunisian firms produce a few vehicle components, like windshields, for European firms. Such activities employ better paid workers, but have not resulted in the spread in these countries of subcontracting, training and transfer of proprietary knowledge as happened with Japanese firms, even though, as in the case of Tunisia, the authorities envisaged subcontracting for European firms as the way for the country raise its income and reduce unemployment over the long run.

Why Japan's firms should behave so differently to the firms of the western high wage countries is a question this work cannot attempt to answer. Any answer would be more concerned with the social and political structures of the high wage countries than the economic circumstances of the low wage countries. For instance, the availability of trained workers was not at first a reason why production was transferred to Malaysia, Thailand or Indonesia. Other countries, like Mexico, Tunisia and Turkey, probably had as many workers with secondary and higher education in the sciences and engineering and their authorities also wanted to attract investment by firms from the nearby high wage countries. When firms from Europe and the US began transferring production to low wage countries in 1961, before Japan and Korea began their collaboration in 1965, it was to lower the labour costs of simple activities. When they did begin to transfer more complex activities, they confined it almost entirely to East Asia, where Japanese firms were already doing the same.

In the mid-1970s, by when the high wage countries had begun to reduce immigration from low wage countries, the construction boom in the Middle East after the oil price rise of 1973 provided a new source of demand for migrant

workers from Pakistan and several other countries. It differed from the immigration of earlier years into the high wage countries of Europe and America in that the immigrants had virtually no prospect of becoming citizens of the places where they worked and only trained workers and those who established businesses could become permanent or long term residents. Untrained workers were almost always limited to contracts of a few months at a time.

For the Pakistani authorities, who kept to their customary belief that economic success was to be measured by the amount of foreign exchange obtained and who took no account of the cost to the economy, it seemed clear from the start that the country should "export" as much labour as possible, and they did what they could to foster this export. Soon the country lost many of its ablest engineers, technicians and other skilled workers. Orthodox economists considered these exports of labour to be a desirable complement to the international mobility of capital and freedom of trade, even if the part of the population that benefited directly was small.

If the authorities believe that sending workers abroad is desirable, it becomes impossible to prevent the emigration of scientists and engineers. They cannot easily help low paid workers to find work outside the country and forbid qualified people from doing the same. At one stage, in 1975, they made an ill conceived attempt to do that, but had to abandon it in a week. Scientists and engineers find it easier than untrained workers to emigrate permanently, for the high wage countries more readily offer them residence and work permits. It is the same for all low wage countries. For instance, roughly 80 per cent of Koreans studying natural sciences or engineering in US universities in the 1960s and 1970s stayed. Korea coped because its own education system was big enough and its firms later got some benefit from the knowledge and experience these scientists and engineers had acquired working in US firms. But for Pakistan it meant scarcities of trained workers, whose costs of training it had borne and who were acquired by high wage countries virtually costlessly.

If trained workers can emigrate easily, this dependence, both psychological and economic, on finding work outside the country, becomes self-perpetuating. As trained workers leave for better pay production becomes inefficient. Even the production of simple goods on an industrial scale needs some trained workers and so do supplying untradables, such as electric power, water, sanitation and road maintenance. Besides, workshop production also loses skilled manual workers, such as mechanics, carpenters, masons and tailors. The administrative system deteriorates, too, as it loses competent and experienced staff. Since the education system cannot replace these people quickly, investment in new production capacity is deterred and, in the extreme, can lead to existing capacity not being maintained or not being replaced as it wears out. Profit is not reinvested but transferred out of

the country legally and at virtually no cost in the expectation that the currency will depreciate. Remittances from Pakistanis abroad become a substitute for domestic production and the economy's dependence on them grows with time, unless the authorities deliberately try to reduce it.

Animosity to the West

Most of the population does not benefit directly from employment abroad, which, however hard the conditions may be, accentuates the meagreness of the prospects at home. Resentment and a feeling of injustice have become widespread. An unskilled worker in a town can see how much better a segment of the population lives and is also aware that in other countries that had been poor, perhaps poorer than Pakistan, for example, Korea, China, India, Malaysia and Thailand, large parts of the population are becoming better off. Peasants in the remotest villages are aware of these things, for they have relatives in towns.

For the mass of the population for whom the present and future are bleak it is natural to blame those they believe to determine economic policy, not just the country's own political or administrative powers, but the foreigners as well who are seen to bring their policy advice along with their economic aid. Institutions like the World Bank and the IMF and the officials and experts of the countries providing bilateral aid are seen to be giving the advice. Most of the politicians and officials who make the main economic decisions are associated with them; they make their public statements in western languages, usually English, using established western economic terms. As a rule they used to wear western clothes, though Pakistani politicians find it prudent not to do so any more. If, therefore, a cause for the failure of the economy to provide for them has to be found, the poor and the middle classes find it in the West; in a country whose people were once among the most pro-western, resentment towards the West has become common.

In the past the poor submitted to the inequities of their societies in the sub-continent with less question. There was no widely held conception of a fairer society or visible example of an Asian country raising the incomes of its poor. They clung to the social framework they had rather than be isolated individuals in a mass with no assurance they would be less poor. When they rebelled against British rule in 1857 they were not rebelling against their landlords and princes but against the changes to their social framework that the British governors of India were trying to bring about in pursuit of Benthamite ideas of progress. Having understood that, the British governors from then on avoided altering the established social and political structures, confining themselves to keeping order and mitigating egregious behaviour in princely states.

Quiescence is a thing of the past in Pakistan, as in most other low wage countries, and social frameworks have changed. Population growth has led to most of the people living in towns, where industry and services provide employment. Old social orders continue more intact in areas where industry provides too few jobs to make a difference, but would not do so had much of the population not left for the towns. Social life in some areas, such as those of the Baluch and Pakhtoon tribes or parts of Sindh, has stayed tribal with the adjunct of relatives in towns, who, in turn, conform to the customs of their places of origin when they return for family events or permanently. But economics and war bring change even there, whilst in the towns people work and compete in industries and services to survive with the desire of being materially better off in the future.

Phantasy alternatives

Many of the organisations helping the poor are religious organisations, in which respect Pakistan and several other low wage countries resemble Europe before the nineteenth century, when it was ruled more or less autocratically by monarchs and nobility, and the little succour there was for the poor, food, medical care and shelter, came mainly from religious institutions. Men and women, humble and noble, often joined or supported these institutions because they wished to help their fellowmen, not for personal advantage, even if the institutions belonged to hierarchies at the heads of which were people with personal ambitions and political goals. Pakistan has no equivalent to the monasteries and convents of Europe, but similar charitable functions are performed there by numerous religious and secular organisations, many of which depend on donations from individuals and organisations, native and foreign, often with their own political and religious motives.

These organisations often improve the lives of people, usually at little cost, but they cannot solve the problem of how to generate continual income growth. They cannot give rise to the production of high wage goods or to autarky. In countries where the state holds out little hope of betterment for the majority of the population, people look for alternatives and the fewer or less palatable realistic alternatives seem, the more likely the desperate are to support unrealistic ones. Communism and some varieties of socialism have been discredited by their failures and any tendency towards them is quickly suppressed. Forms of socialism that have less of the odium of communism because of their distance from it may be more realistic, but are out of vogue. Some manifestations of discontent directed against specific aspects of the economy are proposals that are practical and realistic, though politically they may be neither, but many are movements to return to some idealised or mythical past. These latter occur in the high wage as well as the low wage countries: return to a country's supposed past greatness, to the gold standard, to the founding fathers, to the Caliphate or to the Vedas. They are signs of despair and sources of strife.

APPENDIX

CAPITAL, TECHNICAL PROGRESS AND INTERNATIONAL TRADE

In this work the notion of capital as production capacity is straightforward; it consists of capital goods, i.e. goods that have been produced for the production of goods, and includes capital goods that are, themselves, produced with the use of capital goods, the normal case in an industrial economy. Capital goods are assumed to be operated in combination with labour and to be heterogeneous, meaning they consist of various kinds, each specifically designed for the production of specific goods and not freely adaptable to the production of other goods. Each capital good has a limit to its capacity and will be assumed here to be able to operate for several periods, i.e. to be durable. As an approximation to reality this cannot be objected to.

In the orthodox theory of international trade and in orthodox theory in general, what is called capital is a given, quantifiable mass that is freely adaptable in what it can be used to produce and can be combined with various amounts of labour. This conception of capital can be objected to and was unambiguously shown to be untenable in a prolonged dispute in the 1950s and 1960s. At issue was whether or not the conventional representation of production by a function of quantities of capital, labour and, perhaps, other factors could be justified, and the outcome of the dispute was that it could not, except with restrictions that rendered it useless.

Some account of the dispute is needed here; the way capital is conceived of is not only the crucial difference between the argument of this work and the usual theories of international trade, it is also fundamental to economic theory. Yet orthodox theory continues as though there never had been an objection. Both the occurrence of the dispute and the outcome have been forgotten by all but a few. Perhaps they would be more widely known if more had been done to explain the implications. After all, Sraffa, who first discerned what was at issue, gave his book the subtitle, "Prelude to a Critique of Economic Theory", which, in spite of its apparent modesty, shows his awareness that it concerned the foundations of economic theory.²³³ But the Prelude has not been followed by a systematic critique. Those who took part

²³³ Sraffa, *Production of Commodities by Means of Commodities*.

in the discussions, understood them best and had made the crucial points seem not to have taken the matter further.

Section 1 of this appendix gives a summary of Sraffa's reasoning along with the various attempts to escape or refute the conclusion, none of them valid. Sraffa's original argument was presented for a closed economy with given production techniques. By allowing for international trade the same conclusion is reached by an argument, set out in Chapter 1, that is logically distinct from the original argument but can be regarded as a variant. Allowing for technical progress leads to a third argument, again logically distinct and a variant. This is set out in Section 2 here. Just as Chapter 1 is intended to show that international trade cannot be realistically or plausibly described by production functions and aggregate capital, the discussion of technical progress in Section 2 is intended to show the same for technical progress. Its argument is an elaboration of a point made by Pasinetti²³⁴ in 1959, that technical progress occurs in the production of capital as well as in the production of the final output, a point taken up later by Read²³⁵ and Rymes²³⁶. It follows that the conventional attempts to estimate technical progress as the residual after allowing for increases of factor inputs are fallacious. It also follows that there can be technical progress without there being a residual because it has a cost, which is research and development). Section 3 summarises the conclusions to be drawn.

1. CAPITAL GOODS AND DETERMINACY

Economic theory and depressions

Piero Sraffa seems to have been alone among economists before the Great Depression, if some of Marx's followers are left aside, in believing that there was some flaw common to the various economic theories of his time. According to the various schools the economic system was inherently stable; slumps and booms could occur, but, left to itself, every economy reverted to the full employment of its labour and productive capacity. Thus, Marshall could describe a crash and the argue that recovery would follow quickly, Wicksell could attribute unemployment to the wrong choice of interest rates, Walras could practically ignore the subject and Schumpeter maintain that the disruptions caused by inventions were temporary. Crashes and slumps were familiar to all, but the very term used for their discussion, the "trade cycle", implied that they were transient. Jevons's speculation, that sunspots, which are cyclical, could have something to do with the trade cycle, was in this spirit.

²³⁴ Pasinetti, "On Concepts and Measures of Changes in Productivity."

²³⁵ Read, "The Measure of Total Factor Productivity Appropriate to Wage-Price Guidelines."

²³⁶ Rymes, *On Concepts of Capital and Technical Change*.

Cyclical is a misleading term for the prolonged periods of depression and unemployment before the First World War and more so for the years immediately after. There was good reason to believe that something was wrong with the prevailing theories. It had to be with the theories, for the politicians in power could not be blamed; they had never questioned the orthodoxy before the War and in the new state of affairs after the War they sought and followed the guidance of the experts.

Sraffa seems to have thought for a while that he had found that “something” in the neglect of the implications for perfect competition of increasing returns. In his paper of 1926²³⁷ he argued that supply and demand could not be considered symmetrical in determining prices, as with Marshall’s analogy of the two blades of scissors. Sraffa did not mention the analogy, but argued, first, that the primary determinant of price was the cost of production and, second, that the perfect competition of atomistic producers, none of whom could affect prices, had to be replaced by competition among quasi-monopolies. Increasing returns at the level of the factory were normal and firms producing competing goods could reduce their costs of production by expanding. But more sales by one firm entailed less sales by the others, so, if firms were restrained from expansion, it was by the need to spend on marketing and by the reactions to be expected of the competition. Even with constant returns there was nothing to stop individual atomistic producers from expanding until they could affect prices. Competition consisted, then of producers differentiating and branding their goods to become quasi-monopolists of their brands.

This argument had some success in that it was followed by several works on imperfect or monopolistic competition, notably the books by Robinson²³⁸ and Chamberlin²³⁹. But it did not result in the kind of critique of economic theory Sraffa seems to have had in mind. One reason can be seen from the mathematical representation of production by production functions with labour and capital as arguments. Wicksteed, implicitly assuming constant returns to scale, had shown that payments to factors equal the value of the product if factors are paid their marginal products. Returns to scale, though, are not necessarily constant and when they are not there is a mismatch between factor payments and the value of production. Increasing returns result in the pay exceeding the output and decreasing returns leave a surplus. So, is there to be one theory of prices when returns to scale are constant and another when they are not? Wicksteed’s point shows that something else is wrong with the theories and that the concern with increasing returns is more

²³⁷ Sraffa, “The Laws of Returns under Competitive Conditions.”

²³⁸ Robinson, *The Economics of Imperfect Competition*.

²³⁹ Chamberlin, *The Theory of Monopolistic Competition*.

concern with symptoms than with causes. Sraffa was aware of this, though he did not need to refer to Wicksteed explicitly in his paper.

The concept of capital and indeterminacy

Eventually Sraffa found that “something else” and explained it in his book, “The Production of Commodities by Means of Commodities”. It was the notion of capital as a factor of production. In this paper the discussion is limited to capital as produced goods used for production. Natural resources and investments in exploiting and improving them count as capital and are accommodated in Sraffa’s schema, but were, for simplicity, left out of the disputes and are left out here. Education, too, counts as capital and is left out, because it is imparted to people, and people cannot be treated like machines because of an investment in them. Capital, considered as objects used for production in the production of which the state and firms invest, consists, therefore, of a variety of goods; some, like infrastructure, such as roads and ports, often built without a specific duration in mind; some, like much of the machinery and rolling stock of factories, expected to last a few years, and others, like inventories of raw materials and finished products, turned over in days or weeks.

If the quantity of capital is the value of the produced goods used for production, the common argument that profits and wages are determined by the supply of and demand for capital and labour cannot be used. At any moment the physical capital and its cost are the result of the wage and profit rates of the past and the expectations, when the investment decisions were made, of what those rates and prices would be later, and its economic value depends on what these quantities are expected to be in the future. This is obvious if the physical capital includes durable products that cannot be transferred between uses except at costs that cannot be ignored, what will be termed here as “heterogeneous”. It holds as well if the production process is represented as using only goods that are used up in the process and have no durable capital, provided the goods used in the production process cannot freely be substituted for one another and the process takes time, which is how Sraffa began his exposition.

This contradicts the common argument that an additional unit of capital with the same amount of labour increases output by the marginal product of capital and that this determines the rate of profit. By the same common argument, competition causes the labour used with a given quantity of capital to be paid the value of the output that would be produced by an additional unit of labour and the same amount of capital, labour’s marginal value product. In these arguments it is as though the stock of capital were a malleable quantity and that production were a function of the quantities of capital and labour. The partial derivative of the function with respect to

capital is the marginal product of capital and that with respect to labour is the marginal product of labour. Production alone determines wage and profit rates in a competitive economy, for they adjust to the marginal products of the given quantities of labour and capital so that both are fully employed. Both wage and profit rates, which really determine the value of the capital stock, are supposedly determined by the supplies of labour and capital.

Sraffa demonstrated that, when production is considered alone, there is an indeterminacy: either the wage or the profit rate must be given for the other, hence prices, to be determined. He depicted production as a self-contained system, unchanging through time, each round of production being labour working on produced goods, which are used up in the production process and are, therefore, circulating capital. In this system the price of any good is the sum of the wage cost, the cost of the circulating capital and the profit on it. Each round's cost of the circulating capital is calculated from the prices of the goods from the previous round, which, in a static system, gives the prices of goods as functions of the nominal wage and the profit rate by inverting a matrix. Whatever is taken as the numeraire or standard of value, there is one degree of freedom, meaning that something must be given from outside the system. It can be the real wage, if an unambiguous measure of it exists, and it can be the profit rate, and the two are inversely related lying anywhere between a zero wage and zero profit. As long as both wage and profit enter into the cost of production of capital goods there will be indeterminacy.

Durable, heterogenous capital goods were left out to simplify the reasoning, but can be included by allowing joint production, a device, as Sraffa mentioned, first used by Torrens; each time a capital good is used to produce a given good it becomes the same capital good one period older as a joint output with the good it was used to produce, until it reaches the end of its use. Wage and profit rates are still inversely related, but the mathematics is more complicated, because capital goods of different ages are different goods and each activity has, therefore, correspondingly many outputs. Whether capital goods are durable or not, their prices include both profit and wages. In none of this is there any need to assume constant returns and the process through infinite time can be replaced by a beginning in which the industries are put in place by some other technique of production, which adds to the complication without changing the conclusion.

Disputes about what was meant by capital, how to value it and how it was related to production, prices, wages and profit were nothing new. There is no need to describe here the various schools of thought beyond mentioning that Wicksell had already argued that the profit rate was equal to the marginal product of waiting, which was not equal to the marginal product of capital, and had shown that a change of wage rate could alter both the

prices of any capital goods and the composition of the capital stock. What was new was the indeterminacy, and Sraffa seems to have been the first to spot it, or, at least, to state it explicitly. A letter Sraffa wrote to Joan Robinson in 1936²⁴⁰ seems to show that he had begun to think along these lines by then. In his earlier questioning of constant returns he had had to take the production costs of the firm as given and was perhaps now trying to answer the question as to what determined those costs. He may also have been led to the question of how manufactures were priced when the capital equipment for producing them was not fully used by the discussions of Keynes and his circle as they tried to see how theory could explain prolonged unemployment. But it took twenty-three more years for his book to appear.

Robinson's objections to aggregate capital and production functions

By then the controversy had begun. Joan Robinson's article "The Production Function and the Theory of Capital" had appeared in 1953 and was a criticism of the notion of the production function with capital as one of the arguments. She assumed a closed economy that could use any of a given set of discrete production techniques, along with the usual assumptions of competition, constant returns to scale and an unambiguous measure of the real wage. Comparing stationary states, there is for a given wage a technique that yields the highest rate of profit and each technique, taken alone, gives an inverse relation between wage and profit rates ranging from a zero wage to a zero profit rate. Each technique can be assumed to give the highest profit rate for a range of values of the real wage and at either end of this range, if both wage and the profit rate are positive, some other technique yields the same wage and profit rates as the given technique and then yields the highest profit rate for an adjacent range of values of the real wage.

Robinson posed the question as to how capital was to be valued for each technique. Since capital goods are made with the use of capital goods and take time to make, the value of the capital stock consists of a profit and a wage component. Robinson took the wage as the unit of value, so that the value of a given capital stock is lower when the real wage is higher and the profit rate lower. Other ways of providing a unit of value (normalising prices) are to take a good or a basket of goods as the unit or numeraire.

Regardless of what is chosen as numeraire, when two different techniques yield the same wage and profit rates, the one with a greater value of capital per worker has the greater output per worker and the ratio of the difference of output per head to the difference of capital per head is the rate of profit. These two techniques can be used simultaneously in different

²⁴⁰ Cohen and Harcourt, "Retrospectives: Whatever Happened to the Cambridge Capital Theory Controversies?," 203.

proportions and it follows that, comparing different proportions, the profit rate continues to be the ratio of the differences of output per head to the differences of total capital per head. Usually the technique giving the greater output per head is expected to yield a lower rate of profit, but it does not have to; a technique can yield a higher wage and lower rate of profit than another and have a lower value of capital per head (capital intensity reversal). There is also the possibility that a technique yields the highest rates of profit for one range of real wages and does so again for another range, with other techniques being used in between, what is referred to as “reswitching”. If the wage is the unit of value, the function relating output per head to capital per head is not as usually depicted, a curve with a positive but decreasing slope, even if techniques giving more output per head are assumed to have more capital per head, for, in the range that a given technique is used, as the real wage rises and the profit rate falls the value of the capital stock falls.

From this Robinson argued that ‘the comparison between equilibrium positions with different factor ratios cannot be used to analyse changes in the factor ratio taking place through time, and it is impossible to discuss changes (as opposed to differences) in neo-classical terms’²⁴¹ and that it does not follow, as in the neo-classical doctrine, ‘that the level of wages determines the amount of employment, and that, when unemployment occurs, workers (unless frustrated by the misguided policy of trade unions) offer themselves at a lower real wage rate than that ruling, and go on doing so till all are employed’, even though, given the techniques of production available and ‘the quantity of capital (in terms of product), there is one value of the wage rate which is compatible with full employment of any given labour force’²⁴².

Controversy over capital

After that the controversy concerned two questions, though they were often discussed together. One, the narrower one, was if there were ways around the problems of quantifying or aggregating capital. The other, broader controversy was whether economic theory would have to dispense with production functions and aggregate capital and what the implications could be, but did not extend to technical progress or international trade.

Several efforts to qualify Robinson’s argument followed, the only logically sound one being that of Champernowne, which appeared as a “Comment” along with Robinson’s article.²⁴³ Champernowne showed that a measure of the quantity of capital and a production function with the desired properties of partial derivatives equal to wage and profit rates can be

²⁴¹ Robinson, “The Production Function and the Theory of Capital,” 100.

²⁴² Robinson, 100.

²⁴³ Champernowne, “The Production Function and the Theory of Capital: Acomment.”

constructed by comparing stationary states and using a chain index. He assumed a single consumption good and constructed his index from the value, as given by Robinson's method, of the capital per worker of each technique at the lowest wage of the range in which it can be used and obtained his index from the ratio of this value of capital for each technique to the value of the preceding technique when all techniques used are ranked in order of rising wages and one technique is chosen as the base. When two techniques can be used at the same wage and profit rates they can be used simultaneously in any proportion and the amount of capital is the sum of the amounts of capital per worker multiplied by the number of workers used with each technique. Champernowne gave examples to show that capital intensity reversal and reswitching are possible and that the latter must be excluded by assumption for a consistent measure of the quantity of capital.

Champernowne's chain index does not purport to be usable for actual changes between different steady states; it only compares one steady state with another. Hence, it cannot be used as a production function as commonly done to describe actual changes or to explain cause and effect. It gives the appearance of a production function, as long as there is no reswitching, and has, what Champernowne, himself, and some others considered an advantage, the property that a given stock of capital has the same value whatever the wage and profit rates, again provided there is no reswitching. This last property is not necessarily an advantage; it seems to relate capital as a quantity to the physical stock, whereas profit maximisation is in theory related to the cost.

Later, in 1955, Solow posed the question as an index number problem; how to aggregate the services of different types of capital into one index.²⁴⁴ This was to misstate the problem, as Robinson pointed out.²⁴⁵ Her criticism of the notion of aggregate capital began with the observation that no unit for measuring the quantity of capital had been specified, except value, which depends on prices and, therefore, on the profit and wage rates it is supposed to determine. The question is about how much capital there is in a capital good and presenting it as an index number problem ignores that.

Robinson's point is elaborated on here because it is sometimes misunderstood. Solow presented his index number at first as the question, what are the conditions under which two or more variables in a function can be combined as one variable without altering the function? Solow referred to a theorem by Leontief that gives the necessary and sufficient conditions, which are that the partial derivatives of the function with respect to these

²⁴⁴ Solow, "The Production Function and the Theory of Capital."

²⁴⁵ Robinson, "The Production Function and the Theory of Capital – a Reply."

variables should be independent of any other variables, a condition too restrictive to be considered to occur normally.

Solow's procedure was not what would have been expected; it was to assume a production function with, not capital goods, but two or more types of capital services. Flows of service of capital goods are presumably quantities independent of prices and Solow's production function represents production as a flow coming from flows. Solow may seem to avoid capital goods, but, then, what restricts the size of the flows? Presumably it is the stocks of capital goods, each of which has a capacity that limits the service it can provide in a period. For example, the services can be provided in fixed proportions to the capacities of the capital goods, say to full capacity, in which case the capacities are known from the flows of services. A production function with the services of capital goods, rather than the goods, themselves, as variables has, therefore, to be accompanied by information on the capacities of the capital goods and the proportions of the capacities used. Under competitive conditions the marginal value product of a service equals the price of that service. In equilibrium the price of a service yields the profit on the associated capital good, of which the capacity is known, so that the cost of the capital good can be calculated from the prices of services, themselves calculated from the prices of the goods produced. Or, the prices of the services of capital goods and the goods produced can be calculated from the rate of profit and the values of the stocks of capital goods. Solow does not explain how the rate of profit is determined by a production function that only has services as inputs.

Solow does not eschew capital in this paper; he says earlier, 'For many purposes it is remarkably useful to assume that there exists only one physical commodity which can either be consumed or used as capital in the production of more of itself. Then Q and C are measured in the *same* units except that Q is a flow and C is a stock.'²⁴⁶ Q is the output of the production function and C capital, which he says is measured in "unambiguous physical units", though what the unit is remains unstated. If capital goods are put in place of services in Solow's argument above, each type of capital good is a variable in his production function. Assuming indivisibilities and divergence from constant returns to scale away, there can be blast furnaces of all sizes and there is a measure that indicates the amount of blast furnace, the unit of measurement being a standardised blast furnace. The same applies to a bridge over a river in a particular place, winches, looms and lathes of various kinds in factories and all other forms of capital stock. Complementarities among items would have to be allowed for; more blast furnace requires more

²⁴⁶ Solow, "The Production Function and the Theory of Capital," 101. Italics in the original.

hopper, more storage space for inputs and outputs, more equipment for conveyance, etc. which, together, give more steel mill.

Solow's argument seems to lead to the conclusion that, with a stretch of the imagination, the index number problem can be avoided by making all capital goods arguments of macroeconomic production functions. But the same index number problem makes aggregation of separate production units impossible except under special conditions. As explained by Felipe and Fisher²⁴⁷, the mathematical restrictions on the production functions of the individual production units needed to allow the units to be aggregated into a production function in which the endowments of factors are the sums of the endowments of the units are too restrictive to be thought of as likely to be met or to be construed as representing a genuine economy. Assuming factors to be malleable, therefore, prevents the use of production functions for the whole economy if the same factors are used by the different production units, which they presumably are in different proportions. In contrast, production units can be aggregated if, instead of malleable factors, the capital goods are conceived of as heterogeneous manufactures, as long as indivisibilities and divergences from constant returns can be assumed away.

Movements along Solow's production function are assumed to incur no costs, as if capital stock were malleable. This is standard practice going back to J. B. Clark, at least, and Solow's assessment of it is: 'The kernel of useful truth in' Clark's 'picture of capital as a kind of jelly that transforms itself over time is that indeed, over time something like this does happen as capital goods wear out and are replaced by different capital goods'²⁴⁸ Since the capital goods that replace those worn out have to be made and have a cost, the simile of a jelly, as if the same goods were somehow transformed at no cost, is a bad one. Installing new capital goods is an investment separate from the investment in the goods wearing out and does not necessarily occur.

Various ways of ascribing a quantity other than value to capital were devised, none cogent. Solow, himself, presented one model in which machines are made by labour alone and another in which capital goods are made from a mass of putty, which, once formed into machines, is unalterable, and Swan proposed that machines be made from a set of elementary components, each of which embodies a quantity of capital. In these no capital is used for forming capital goods, so that they are ways of having a quantity of capital without profit in its cost of production and, since different capital goods can be made from the same labour, putty or components without other costs, they are various forms of malleable capital.

²⁴⁷ Felipe and Fisher, "Aggregation in Production Functions: What Applied Economists Should Know."

²⁴⁸ Solow, *Capital Theory and the Rate of Return*, 27.

Somewhat later Samuelson proposed the notion of “surrogate capital”, a quantity that could be deduced from a model he devised and yielded a production function with the properties desired.²⁴⁹ His model had a single consumption good and several types of durable machines, each of which could, in combination with a fixed amount of labour, produce the consumption good or machines of the same type. The highest rate of profit, given the real wage, (or *vice versa*) that could be obtained in each stationary state defined the “factor price frontier”, which could be approximated by a production function with surrogate capital. There would be no need for approximation if the machine types were a continuum. But he also pointed out in the same paper that Garegnani had shown that the condition for surrogate capital is that each type of machine use the same amount of labour for producing the consumption good as it uses for making the machine. Garegnani showed later that the condition is necessary and sufficient and, in effect, results in a single good economy, for, in stationary states, the machines are just intermediate steps in the production of the consumption good by itself and labour.²⁵⁰

Finally, the first question of the controversy was considered settled after Levhari had claimed to have proved the assertion that, if production can be described by an indecomposable matrix, meaning that every good enters directly or indirectly into the production of every good, reswitching cannot occur for the whole matrix.²⁵¹ Pasinetti showed that there was a mistake in the proof and soon several counterexamples were given to show the assertion to be wrong.²⁵² Champernowne’s production function and the associated measure of capital can be used to compare steady states if it is known there is no reswitching, but the common practice of using production functions with capital as an aggregate quantity to describe change over time has no justification.

Attempts to remove the indeterminacy: fallacies with capital and production functions

With the first question settled, the second question of the implications for economic theory became, what was the critique to follow the prelude? Several economists have claimed that no critique need follow, for, beyond showing the need for care with some simplifications, the point at issue does not affect the validity of neoclassical economic theory or of the applicability of marginal costs. They have made various attempts to justify the claim, all either fallacious or mathematical abstractions remote from reality.

²⁴⁹ Samuelson, “Parable and Realism in Capital Theory: The Surrogate Production Function.”

²⁵⁰ Garegnani, “Heterogeneous Capital, the Production Function and the Theory of Distribution.”

²⁵¹ Levhari, “A Nonsubstitution Theorem and Switching of Techniques.”

²⁵² Pasinetti, “Switches of Technique and the ‘Rate of Return’ in Capital Theory.”

APPENDIX

First are arguments intended to remove indeterminacy, but which proceed by assuming what is in dispute. Of these the earliest seems to have been Solow's attempt, in his De Vries lectures of 1963, to show that conclusions drawn from assuming malleable capital can be reached without that assumption and that a rate of return on capital can be calculated from a reduction of consumption in one period to allow an increase in capital goods so as to yield an increase of consumption the next period. Beginning from an efficient allocation, so that the production of no good could be increased without reducing the production of some other good, he considered a reduction of consumption and asserted 'Because all the allocations considered are efficient, those which produce less consumption must also produce more of at least some kinds of capital goods'.²⁵³ Here the logic is faulty. He did not and could not assume all allocations to be efficient. Simply reducing consumption results in an inefficient allocation if the capital goods producing the consumption cannot be used for producing capital goods. Solow is either tacitly assuming that capital is malleable, though he states in the same work the assumption is 'obviously absurd' and is not needed for 'neo-classical capital theory'²⁵⁴, or he is assuming that the labour released by reducing consumption can make capital goods without using capital goods or that the existing stock of capital goods was not being used to capacity and the initial state was inefficient.

Solow purports to give an example that does not have malleable capital but shows how 'in competitive equilibrium the rate of interest must equal the rate of return on investment'²⁵⁵, the example being a model of Worswick, in which capital goods are assumed to be made with labour alone. He claims the assumption is a simplification 'without being in the least necessary'²⁵⁶, though, as already mentioned, it removes profit from the cost of capital and allows capital to be quantified as an amount of labour. Böhm-Bawerk already knew better.

Later, in 1982, Hahn tried to argue, firstly, that Sraffa's static model is merely a special case of more general neoclassical theory and adds nothing that cannot be obtained from the latter and, secondly, that Sraffa was wrong to assume a uniform rate of profit in a static system and thus to restrict unduly what should be an intertemporal equilibrium.²⁵⁷ Hahn points out that Sraffa's main argument is made with a model that has one technique of production and is adapted to several techniques only towards the end of the book, whereas

²⁵³ Solow, *Capital Theory and the Rate of Return*, 18.

²⁵⁴ Solow, 26.

²⁵⁵ Solow, 33.

²⁵⁶ Solow, 33.

²⁵⁷ Hahn, "The Neo-Ricardians."

neoclassical theory allows for an unlimited number of techniques, and he asserts that, correctly formulated, the standard marginal relations still hold.

For his first argument Hahn used a model with two goods and a choice of production activities, each activity using both goods and labour to produce one of the goods. Different techniques of production use the two goods and labour in different proportions. Taking at first a single period and assuming that the production techniques can be represented by differentiable functions of the inputs of the goods and labour, there is a consistent set of equations relating the marginal products to the prices and wage, given the rate of profit.

From this Hahn concludes that, 'under assumptions no more stringent than Sraffa's'²⁵⁸, where 'you cannot get more neoclassical than differentiable' functions representing production techniques, the rate of profits is still not determined though 'every possible marginal product had been used'²⁵⁹. One more equation is needed for complete determinacy, that is to determine the profit rate, just as with Sraffa, and that implies 'the meaninglessness of a sentence like: 'the marginal product of labour determines the real wage''²⁶⁰ since all the equations must be solved simultaneously.

One of Hahn's assumptions, however, is more stringent than Sraffa's and that is that the same goods are used in different combinations to produce the same goods. If one of them were to be called "capital" and assumed not to be consumed, the assumption would be seen to be that of the production function with malleable capital that Sraffa had criticised plus a second input. Hahn avoids the term "production functions" when referring to the functions representing the production techniques, though that is what they are. This assumption also provides the equation needed to determine the rate of profit by the standard neoclassical argument that competition results in the full employment of labour; the wage falls if there is unemployment and rises if there is a shortage of labour until it and the marginal product of labour are equal. In Hahn's model, therefore, the marginal product of labour does determine the real wage, even if Hahn, himself, omits this piece of neoclassical theory.

So, when Hahn refers to the problem of finding a measure of the capital of his model, assuming both goods are inputs into their own production, he merely restates Solow's index number problem. To obtain his measure of capital he modifies his model for a two period variant with an initial endowment of the two goods, which, since it is used for production, can be thought of as capital and a function of it 'may be thought of as a measure of

²⁵⁸ Hahn, 360.

²⁵⁹ Hahn, 361. Italics in the original.

²⁶⁰ Hahn, 361.

capital stock'²⁶¹. But, knowing the value of the function 'would not allow us to 'determine' equilibrium',²⁶² since different combinations of the two goods can give the same value of the function, but different equilibria. Then, other values of the measure of capital stock given by other combinations of the quantities of the goods are possible, which is to say that it is not possible to combine the two goods without altering the production functions of the model, except in special cases.

Nevertheless, Hahn's procedure of assuming that production can be represented as using the same goods as inputs in different proportions is common. Ethier used a similar model for the same purpose.²⁶³ Both Samuelson and Solow described a procedure of supposing a finite number of techniques with no reswitching and then supposing that the number increases in such a way that the set of profit rates at which two techniques are used becomes everywhere dense. They also needed the assumptions that a technique used at a lower profit rate than another has a greater value of capital per head and that the difference of capital per head between two techniques tends to zero if the difference in profit rates tends to zero. Samuelson used this procedure to make his continuously differentiable "surrogate" production function from a beginning with a finite number of techniques. Another of several examples is that of Burmeister.²⁶⁴ Unlike these, Champernowne's production function has the desired properties of partial derivatives with respect to labour and capital being equal to the wage and profit rates respectively, yet does not need that procedure because it is confined to a finite number of techniques.

Less stringent and more realistic is to accept that different techniques for producing the same good use different varieties of capital equipment. If a firm does change its technique for producing a specific good, some capital equipment is changed, with the concomitant costs, though some, like buildings and roads, may not be. Equipment cannot be supposed to consist only of goods that can be combined in different proportions. This is obvious when the production possibilities are depicted as a finite number of activities, which must, then, be discrete. It is obscured by the assumption of continuity, when it becomes harder to imagine that the equipment may be different if the change of wage or profit is infinitesimal.

This point, that different production techniques may have different capital goods, has a counterpart with consumption goods. If it is no longer assumed that there is only one such good, the possibility of different

²⁶¹ Hahn, 361. Footnote.

²⁶² Hahn, 369.

²⁶³ Ethier, "The Theorems of International Trade in Time-Phased Economies."

²⁶⁴ Burmeister, "Wicksell Effects."

compositions of consumption must be allowed for. Along with differences of production techniques there are differences of consumption patterns and the comparisons are as of two countries that can use the same production techniques, but of which the peoples have different preferences. Production techniques can still be ranked by their profit rates, but when two countries with different profit rates have different patterns of consumption and, consequently, different stocks of capital goods, the procedure of combining two techniques in different proportions, which posed no problem with one consumption good, now must cope with differences of people. The counterpart to the combination of different production techniques involves peculiar assumptions about how to combine different peoples.

Attempts to remove the indeterminacy: demand and equilibrium

Second, there are the arguments that the indeterminacy between wage and profit can be removed by bringing in demand. If demand determines prices, it determines wages and profits. This does not contradict Sraffa's argument, which is, that to consider production alone is not enough to determine the economy and which refutes the common, textbook argument using production functions that seems to give determinacy without reference to demand, namely that competition causes wage and profit rates to become equal to their marginal products. But it adds the condition of consistency, for wages and profits also determine demand.

In the simplest case, that of the steady state, the indeterminacy is evident when demand is left out. For a given set of available techniques of production and a given constant population growth rate, there is a range of steady states, each with its own wage and profit rates and technique of production, and there is no reason to select one state rather than another. Demand can be brought in to determine the state by making assumptions about saving, for instance by assuming constant rates of saving out of wages and profits. It is possible for saving not to suffice for equipping all workers as the population grows and old equipment is discarded, even with a constant population, yet for the economy to be in a steady state. Since the gross investment in each period that keeps a given proportion of a constant or steadily growing population employed replaces worn out equipment and equips any additions to the workforce, the technique of production being used, which determines the distribution of income between wages and profits, must be one that gives the right amount of saving. Since returns to scale are constant, any proportion up to that giving full employment meets the condition of consistency, which adds another form of indeterminacy. This second form of indeterminacy can be accepted as a concern of the country's economic authorities, or assumptions can be made about the behaviour of wages and investment in addition to those about saving

rates in an attempt to remove it. In either case the economy moves out of the steady state.

In theory an economy can be in equilibrium, though not in a steady state, the criterion for equilibrium being that no firm or individual gains from change if none of the others change, which implies that all expectations are fulfilled. With capital goods that are durable and heterogeneous changes of prices, capital stock and production become too complicated to be explicitly described. Nonetheless, firms and households know what to expect, for fulfilled expectations means, among other things, that the profits on an investment in equipment discounted over the time that the equipment is used yields the anticipated rate of profit, which implies both a constraint on the changes of prices of output and capital goods and that the changes are foreseen. Competition implies uniformity of prices at any time and, therefore, that the same discount rate has been used for all investments at the time, though that rate can change over time.

This equilibrium is general and intertemporal. Prices can change from period to period and the profit rates they yield for any one period can, therefore, differ from good to good. Hahn argues that, because of this, a uniform rate of profit each period, as assumed by Sraffa, would, at best, be a special case of equilibrium.²⁶⁵ His point is valid and he illustrates it by his model with two malleable goods, now called wheat and barley, and two periods. But that is only to repeat that Sraffa's indeterminacy does not occur in models of intertemporal general equilibrium, which are necessarily determinate.

Models like that of Arrow and Hahn²⁶⁶, which had its origin with Debreu, have assumptions about production and consumption at least as general as those of the steady state and do have such intertemporal general equilibria. Arrow and Hahn assume consumers to have consistent preferences for the various goods over the various periods, whilst the assumptions about production are general enough to accommodate durable, heterogeneous capital goods. Consumers and producers behave rationally, which, for the former, is defined as maximising their welfare over time within the limits of their incomes and, for the latter, maximising profits. That there is an equilibrium is proved using a fixed point or a separating hyperplane theorem, both of which are statements that something exists without indication of how it is to be found. Equilibrium is shown to be efficient in the sense that no individual or firm can be better off without some other individual or firm being worse off.

²⁶⁵ Hahn, "The Neo-Ricardians," 363–64.

²⁶⁶ Arrow and Hahn, *General competitive analysis*.

These are, however, purely mathematical results that leave open the questions as to how they can be applied to the real world and how they can be extended to international trade. In his criticism of Sraffa, Hahn, refers to a similar ‘... crucial and beautiful theorem in neoclassical economics ...’ and states, after listing the conclusions, ‘These results are theorems and they are not at risk.’²⁶⁷ Arrow and Hahn do not assert that their model is realistic; their justification is that it shows that ‘... an economy motivated by individual greed and controlled by a very large number of different agents ...’ does not end in chaos but is ‘... compatible with a coherent disposition of economic resources that could be regarded, in a well-defined sense, as superior to a large class of possible alternative dispositions’²⁶⁸. Taken literally, they seem to imply that the problems of economic development can be solved by starting with anarchy. No need for institutions, government or economic theory, and no need to study economic instability.

More reasonable would be to argue that the impossibility of the requirements is proof to the contrary; that equilibrium in the sense used here cannot possibly occur. Few take such models literally and it may seem superfluous to give reasons as to why they are too far-fetched to be used to draw conclusions about how economies work in practice, especially as neither Hahn nor any other proponents of models of this sort claim these models can describe actual economies. Despite their apparent generality, the assumptions require too much that is impossible, including perfect foresight, perfect forward markets for everything and individual consumers whose preferences ignore birth, death and family. And no model of comparable generality seems to have been formulated for several open economies; they are all confined to the closed economy. These models are also vague; their proponents do not say in what way they are relevant to reality, but seem to imply that they are relevant in some way, which is why it is not superfluous to make some obvious points in what follows.

Their very generality is a source of vagueness. The assumptions about production may be general enough to allow for some deviation from constant returns, externalities and indivisibilities, as well as for some technical progress, but there seems to be no way of telling how much and in what form. Similarly, the assumptions about consumption perhaps allow for the welfare of an individual, say a child, to influence the welfare of another, say a parent, but it is not clear to what extent education and upbringing can be allowed for. Perhaps mathematical research would yield information as to the extent that these and other complications can be accommodated, but it would only be of academic interest and probably uninteresting mathematics.

²⁶⁷ Hahn, “The Neo-Ricardians,” 371.

²⁶⁸ Arrow and Hahn, *General competitive analysis*, vii.

Its complexity alone makes intertemporal general equilibrium impossible. Equilibrium results from the rational behaviour of each household and firm and, at the same time, determines that behaviour by providing the information households and firms need to decide how to act, notably prices, demand and the production possibilities of firms through the future. Each household and firm knows how all the others intend to act or is somehow provided all the data it needs to act consistently with the others. Among the reasons that this is impossible is that any but the simplest optimisation requires an expertise in programming that few households can have and optimisation of this complexity is impossible, even if the data needed are available.

Behaviour when coping with unmanageable complexity and uncertainty is qualitatively different to the rational behaviour of models of general equilibrium. Households must find methods other than intertemporal maximisation of welfare to cope with the complexity and uncertainty of their economic lives. Usually they make budgets. It is what practically all institutions, including those constituting the government of the country, and big firms do. When households seek expert advice on how to manage their incomes and daily finances, they are told to prepare budgets. Even households that do not explicitly plan their expenditures by budgeting can have budgets from habit or from rules they have in mind, and some kinds of behaviour that seem irrational because they do not conform to intertemporal maximisation of welfare may be explicable as attempts to stick to budgets. Not all households can be assumed to use budgets regularly; some are less careful with their spending and others too rich to bother. But they cannot be imagined to optimise their expenditure. Household budgets are easily revised with changes of pay, employment, inflation and so on, in contrast to the budgets of government and many firms and institutions, which must also have budgets because they are answerable to others for the monies with which they are entrusted.

Some economists argue that complexity does not pose a difficulty. They compare the household or firm to the billiard player, who does not need to know Newton's laws of mechanics. As an analogy it is misleading by being vague, since, if it applies to anything, it applies to the knowledge involved, not the execution. It implies that mathematicians are good at billiards because they know the theory. It is also irrelevant, for, whether or not households and firms know economic theory, it is the complexity of the calculations they cannot cope with; though most people can learn to play billiards and improve with practice, general equilibrium implies they make perfect shots straight away. Not even the best player performs perfect shots all the time, and players vary in their abilities and in the time they devote to practice. If the analogy were valid computers would not be needed for the

design of aeroplane wings or bridges, or, to put it differently, this recourse to analogy shows there is no logical argument.

Economics without equilibrium

Presumably the proponents of intertemporal general equilibrium believe actual economies approximate their models, or would do so if left to themselves, for the models would otherwise be no more than riders to theorems in topology and Hahn's references to such a model in arguing against Sraffa would be pointless. But the belief that such models resemble reality is self-contradictory. Expectations are often mistaken and plans of households, firms and governments are almost always modified, if not changed altogether, to match actual outcomes. To say that this is in some way an approximation to an intertemporal general equilibrium is to say that the equilibrium is not determinate. Some objections to the argument that the equilibrium would occur if the economy were left to itself have already been given.

Models devised for forecasting an economy's behaviour are, therefore, necessarily conjectural; they must rely on assumptions as to how households and firms behave and on simplified descriptions of production. They are not simplifications of or approximations to general equilibrium models with durable, heterogeneous capital goods applied to available data, nor can they reproduce the optimisation by households and firms that general equilibrium models presuppose. As an example, no satisfactory way of predicting saving has yet been found. In theory a forecasting model should depict production as flows of current inputs (raw materials and intermediate goods) into production processes constrained by capital stocks to produce consumption goods, raw materials, intermediate goods and capital goods. It should, therefore, have two input-output matrices, one for current inputs and the other for the capital stock, which Schwartz seems to have been the first and practically the only one to propose.²⁶⁹

No model of this kind with two matrices seems to be in use for economic forecasting. One reason may be that such models are more onerous to put together than conventional ones. Getting the data for the input-output matrices should not be as difficult nowadays as it was some time ago, but investment and the corresponding production need to be specified to the same level of disaggregation and, so, forecasts must rely on information about their intentions that firms, domestic and foreign, and government agencies provide. Production functions have the apparent advantage of not needing so much information; once the type of the function has been chosen, it takes relatively little to fit its few parameters. They also simplify investment, which becomes a quantity of capital. It is the hope that convenience can substitute for reality.

²⁶⁹ Schwartz, *Lectures on the Mathematical Method in Analytical Economics*.

Pragmatism and production functions

Considering the quantities of data that are routinely collected by various agencies, including central banks, this does not seem enough to explain the prevalence of models with production functions. Thus, all central banks that have models for their countries' economies use production functions, mostly CES²⁷⁰ functions, though the German central bank and some others use Cobb-Douglas functions. A more likely reason for the prevalence is that the people who make the models are not aware of the objections to production functions. Forecasts are routinely tested by the outcomes and experience shows the models used for them to be unreliable, so it would be expected that the models would be rejected, as they would be in the natural sciences. But, in practice, the standard of comparison for economic models is the accuracy of the forecasts of other models and, since they are all more or less alike, they all give equally unreliable results. As a matter of experience, a model that gave a more accurate result than others in one period cannot be relied on to give an equally good result the next. Economies are hard to predict, even in the short run, and it is accepted that those who make and use the models are not to be held accountable for the accuracy of their forecasts. Instead of the models being rejected, forecasts are adjusted as the actual outcomes become known.

Economic models for forecasting only became common as computers made their calculations feasible, but fitting Cobb-Douglas production functions to national output has been going on for longer.²⁷¹ In these early exercises the functions fitted the data well and, since it seemed that some kinds of theoretical and empirical work would be unmanageably complicated if the heterogeneity and durability of capital goods were allowed for, some argued in favour of using production functions and treating capital as though it were a malleable substance on the grounds of being pragmatic and practical. In his De Vries lectures Solow likened his empirical work on the 'social return on investment' to '... what the inveterate gambler said about the dishonest roulette wheel, 'I know the wheel is crooked, but it's the only game in town'.²⁷²

The gambler's fate is assured, as Brown had already shown in 1957.²⁷³ Brown showed that the fit of the production function to the data was good because of the regularities of the data, notably similar growth rates of factors and similar rates of pay and returns on investment across the economy. He also agreed with others who had pointed out that it was improbable that 'one

²⁷⁰ Constant elasticity of substitution.

²⁷¹ Brown, E. H. Phelps, "The Meaning of the Fitted Cobb-Douglas Function," 546.

²⁷² Solow, *Capital Theory and the Rate of Return*, 68.

²⁷³ Brown, E. H. Phelps, "The Meaning of the Fitted Cobb-Douglas Function."

unchanging production function should fit a growing, changing economy over a run of years' and remarked that attempts to add time as a variable to the production functions had given results that were not acceptable, rather than better.²⁷⁴

2. CAPITAL GOODS AND TECHNICAL PROGRESS

Defining technical progress.

For present purposes and without attempting to give a precise definition, technical progress is taken in the broadest sense allowed by the assumptions that there is at any time a set of known goods that can be made in one or more versions and a set of known production techniques for making them, and that there are several countries with different nominal wages; it is the addition of new goods or of new versions of existing ones and the addition of new production techniques that can yield positive profit for some combination of nominal wages among the different countries whenever potential producers have the requisite training and knowledge. If a new production technique cannot be profitable with any consistent set of wage and profit rates in the various countries, it is not to count as progress. It is taken for granted that countries do not differ in how the requisite training and knowledge can be imparted, though the imparting may not occur, nor in how well any technique can operate. Goods can be for consumption, can be capital goods and can be intermediate goods, including raw materials, and they can be durable or used once only. New production techniques can be expected to involve new types of capital or intermediate goods, but allowance is made for the possibility that they do not. In this, the common sense view, technical progress is equivalent to acquisition of knowledge by which new and better production techniques and products are made possible and allows for such knowledge to be generated deliberately and at a cost through R&D, though not necessarily a cost that is recovered.

Skill acquired from practice, as with Adam Smith's pin maker, does not count as technical progress; it is not a new technique, but is confined to the individual and is an improvement of efficiency that must be supposed to have a limit. Arrow argues that proficiency can improve with experience or practice, "learning by doing", which seems to be the same as with Smith. But he implicitly assumes that experience is not confined to individuals and results in improvements of capital, newer capital requiring fewer workers per unit than older capital. It is the reverse of Smith's pin maker and his own example of the Horndal steel mill, examples to show that existing equipment can be expected to be more efficiently operated because of experience. Rather, it is an improvement with practice of design, not production. There

²⁷⁴ Brown, E. H. Phelps, 548–50.

is also no limit to the improvement possible. Hence, learning by doing in Arrow's model does count as technical progress.

Technical progress and product differentiation

Most goods have several characteristics that users like or dislike and there can be several versions with different combinations of these characteristics at any time. Lancaster had already proposed associating a combination of characteristics with each good in 1966.²⁷⁵ For present purposes such goods are termed differentiated, whilst those that are not differentiated are referred to as homogeneous. In some cases competing firms may make the same version of a differentiated good and compete solely on price. But often versions of the good differ because of design and the designs are protected by intellectual property rights, like patents, copyright and trademarks, which allow each owner of a design to prevent others from copying it or imitating it too closely. Furniture, clothing and houses, for example, are mostly designed for both practical and aesthetic reasons and the designs are normally so protected. Each version that is protected is a quasi-monopoly of the producer. None of the different versions of a good need be unambiguously better or worse than any other, for one may be preferable to another in some respects but not in others. Or one may be better than another but costlier. Consumers buy according to income, taste and convenience. Competition among producers of differentiated goods of which the designs of the various versions are protected consists of designing versions with combinations of characteristics that do not infringe on the intellectual property rights of others and, as quasi-monopolies, fetch high enough prices to be profitable.

This diversity of competing versions at any time does not, in theory, carry over to capital goods. Different versions of a consumption good, including more expensive, better versions and cheaper, inferior versions, can be sold at the same time because consumers differ in their preferences and incomes. In contrast, firms all have the same preferences, profit maximisation, and their incomes do not matter if returns to scale are constant. So, if firms with the same wage and profit rates choose different versions of the same capital or intermediate goods to produce the same good, they do so because the different versions of the good they produce require different types of equipment. In reality the capital goods for producing the same good are differentiated for other reasons as well, among them returns to scale not being constant, externalities and physical circumstances, including costs of untradables, considerations that have to be mentioned but are not pursued here.

²⁷⁵ Lancaster, "A New Approach to Consumer Theory."

Technical progress is obviously a source of differentiation when it results in new goods or new versions of old ones. It can result in a new version being better than an old one because it increases some desirable characteristics and lessens some undesirable ones, or because it adds new, desirable characteristics and eliminates old, undesirable ones. Motor cars illustrate this; cars can differ in thousands of ways and technical progress constantly results in improvements leading to new versions, though they are all versions of the same good. But the distinction between a new good and a new version of an existing good is to a great extent arbitrary. Some components of a car may be considered new goods, as, for example the automatic transmission and anti-blocking system in their times, though the cars that included them were not considered new goods. Most people would think of the locomotive, which was an adaptation of the principle of the steam engine from mining, to have been a new good. But did the use of a similar engine in a ship mean the steamship was a new good or a new version of a ship? If the steamship was a new good, was it partly because it used a propeller or a paddle wheel, which sailing vessels did not? The incandescent lightbulb has been followed by fluorescent lights, mercury bulbs and light emitting diodes, all of which work on different principles. But are they different versions of the same good or different goods?

For the purposes of the discussion here there is no need for definite criteria. But the distinction has to be made because in the endogenous growth models discussed later there is a distinction in the utility and production functions. In these models a new consumption good is added to the other goods in a utility function, so that the individual consumer may be consuming that good at the same time as the older ones, whereas the newer, better version of an existing consumption good is chosen to the exclusion of the older version. Similarly for capital and intermediate goods used in production; new ones are added to the production functions and new versions displace old versions.

Types of technical progress

However the distinction between new versions and new goods is made, technical progress can be classified into four forms. Two are the improvement of existing goods and the invention of new goods, whether for consumption or production. A third is the improvement of processes of production of the same goods using the same labour, goods and factors, provided it is independent of the experience of the workers operating the processes. If it is not, it is Smith's acquisition of skill, which can be supposed to have a limit. A good that is improved is necessarily differentiated, for, at the least, there are a new and old version. A new production process that has new or improved goods as inputs or output is subsumed under the first

two forms. The fourth form is an increase of utility from the same consumption using knowledge generated by R&D and for which people are willing to pay. It is only mentioned for the sake of completeness since there seem to be no models of the kind and it would not be manifested in the national accounts. For brevity, most of the discussion in the following is confined to the first two forms. The third form, being less important, is referred to only when necessary and the fourth can be ignored.

It can be supposed that technical progress increases the welfare attainable. More technical progress is better than less, other things being equal, but not all of it is relevant to any particular economy. Technical progress in the general sense used here is progress of technical possibilities and is to be distinguished from the application to a specific economy. A new production technique may be used in one economy but not in another because it is unsuited to the wage and profit rates of the second. It is, in principle, possible for a new technique not to be suitable to any economy because of the existing set of wage and profit rates, though it would be for some other set and, therefore, constitutes technical progress, though not relevant under the circumstances.

Technical progress is usually not applied immediately everywhere it can be. Durable goods are often not displaced as soon as there are new versions and every economy has stocks of various ages. Families do not as a rule change their television sets the moment an improved set is available, better ways of conserving energy in homes spread gradually and most of the housing stock in any country is technically out of date. Improvements of production techniques usually involve new types of capital equipment and the rates at which they replace equipment in use depend on, among other things, how much of an improvement they are and the ages of what is in use. As the new versions and new goods are adopted yet newer ones come into being. Hence, the rate at which technical progress actually takes place in an economy depends partly on the rates at which improvements and inventions occur and partly on how fast they displace old goods, both consumption and capital goods, and differs from one economy to another according to the compositions of their stocks.

Several questions arise; among them, what are the causes or explanations of technical progress, can they be influenced and, if so, how, what are the benefits to other countries of technical progress in a given country, what are the gains and costs of faster adoption of newer versions of consumption and capital goods, how can the effects of technical progress on welfare be assessed and can the technical progress of an economy be measured? Only the last question can be answered here; the technical progress of an economy as a whole cannot be quantified. For the rest, it is the purpose of the

following to show that the questions cannot be realistically discussed, let alone answered, if production is represented by functions of labour and capital with, perhaps, other factors, or of labour alone.

Using production functions to represent the aggregate production of an economy eliminates the first two forms of technical progress for consumption goods. Output is represented as a single quantity, which means its identity and composition do not change and, therefore, that technical progress in the form of new and improved goods cannot be allowed for. There is as yet no objective way of allowing for these types of technical progress in the usual measures of aggregate output, such as GDP. Sometimes notional adjustments are made to allow for improvements of consumption goods, but they are not objective. It would make no difference if output were to consist of several goods, each with its production function, since each would still be a single quantity. The same is true for capital in models with one good or with a fixed set of types of capital all of which are equally available. Pasinetti's point referred to earlier is that capital goods also change and, even if a capital good does not change, the capital goods for making it may. To take the capital in a production function as given is to leave out this part of technical progress.

Technical progress as a residual

If neither the technical progress in consumption goods nor that in capital goods is to be considered, only the third form is left; there is no alternative to arguing that, apart from technical progress "embodied" in goods, that is the first two forms, there is "disembodied" technical progress. That is how the attempts to estimate technical progress of Solow, Denison, Jorgenson and Griliches and others must be understood. Even if it is conceded that output per head can increase without change of physical equipment and can do so without limit, the assertion that disembodied forms of technical progress are so much greater than the embodied forms that the latter can be neglected, apart from being hard to imagine, requires both empirical evidence and some description of how that technical progress takes place.

As the following brief survey shows, these attempts at estimates of technical progress are devoid of any explanation or description of the technical progress, itself, but identify it with that part of the increase of output that is not explained by increases of factors. In other words, it is a residual. Then what is the evidence that it is not, as Abramovitz stated regarding Denison's 1962 estimate of the residual, termed the Advance of Knowledge, '...the grand legatee of all the errors of estimate embodied in the measures of national product, of inputs conventional and otherwise, and of the economies of scale and other factors classified under productivity

growth.²⁷⁶ If different methods had yielded similar results there could have been some grounds for supposing that they were measures of the same thing. Instead, the results of the different estimates discussed below were sufficiently far apart to cause disagreement over how the statistical data should have been used, although the method was the same, namely using prices as indicators of marginal products.

Solow was the first to try to estimate technical progress from an aggregate production function by equating it to the residual. In his article published in 1957 he excluded agriculture from his estimate and assumed labour and capital to be the only factors, so that their incomes as given by the national accounts add to the value of the output.²⁷⁷ He assumed factors were paid their marginal products, which allowed the further assumption that the production function was linearly homogeneous in the two factors. Technical progress was assumed to be Harrod neutral, meaning that the marginal product of capital does not change if the ratio of capital to output does not change. Solow claimed that the data allowed this assumption and assumed a production function in the form of a static production function multiplied by a factor representing technical progress that grew with time, which gives Harrod neutral technical progress. It follows that the rate of technical progress, which is the rate of change of its factor, is obtained by deducting the increase of capital per head multiplied by the share of capital in income from the increase of output per head.

From this Solow estimated that technical progress accounted for seven eighths of economic growth from 1909 to 1949. The scatter diagram showed a positive, almost linear relation between output per head divided by the factor for technical progress and capital per head, except for the last seven years, which lay on a parallel but separate line and had to be left out for lack of explanation for the difference. In other words, if output per head is adjusted to allow for technical progress, it increases in proportion to capital per head. Solow fitted five formulae for production functions to the data and obtained high correlations for all of them, the highest (0.9996) being obtained by a Cobb-Douglas function.

Later Solow did introduce embodied technical progress in two models, but their purposes not being to estimate or explain it, they need only be briefly described. In the one the capital created in one period is more productive than the capital of the previous period in the sense that the production function of the former is equal to that of the latter multiplied by a factor representing technical progress.²⁷⁸ Total output at any time is the

²⁷⁶ Abramovitz, "Economic Growth in the United States," 775.

²⁷⁷ Solow, "Technical Change and the Aggregate Production Function."

²⁷⁸ Solow, "Technical Progress, Capital Formation, and Economic Growth."

sum of the outputs of the stocks of capital of that and each earlier period, with due allowance for depreciation. Solow's purpose was to discuss the relationship between unemployment and output taking the capital of each period as given and assuming various rates of technical progress, which he did not try to estimate. His other model was intended to show that the marginal products of labour and capital can be defined and are equal to the wage and profit rates even when capital goods are durable and heterogeneous.²⁷⁹ For this he assumed that machines for making the consumption good are handmade and that there are different types, those using less labour to produce a unit of the consumption good requiring more labour to be made. Assuming that machines are handmade allowed him to avoid the complications of machines making machines, etc., but Solow added, '... I do not believe that any matter of principle is involved.'²⁸⁰ He did not mention that he also avoided the complication of profit entering into the cost of machines and that his conclusions depended on that.

Jorgenson and Griliches used the same theory as Solow, but in their article of 1967 they came to virtually the opposite of Solow's result, namely that little of the increase of output per head could be attributed to the residual.²⁸¹ Where Solow had concluded that seven eighths of the increase of output over the years 1909 to 1942 came from technical progress, they estimated that 79.8 per cent of the increase of output from 1953 to 1965 was explained by the growth of inputs and that the residual had grown at 0.72 per cent a year.²⁸² They attributed the differences to different statistical procedures. Also, they did not, as did Solow, estimate an explicit production function, but assumed, instead, that the relation between outputs and inputs could be represented by an implicit function without needing to specify its form. Then, when competition is perfect and returns to scale are constant, so that the prices of inputs and outputs are inversely related to the relevant partial derivatives of the production function, the difference between the price weighted indices of outputs and inputs is zero. If, therefore, the data show that the output index has increased more than the index for inputs from one period to the next, the implicit function must have changed and the difference between the increases is the residual. Jorgenson and Griliches used Divisia index numbers over several years and called the residual "total factor productivity".

In turn Denison²⁸³ criticised their procedures, though the differences of procedure need not be described here. Jorgenson and Griliches revised their

²⁷⁹ Solow, "Substitution and Fixed Proportions in the Theory of Capital."

²⁸⁰ Solow, 207.

²⁸¹ Jorgenson and Griliches, "The Explanation of Productivity Change."

²⁸² Jorgenson and Griliches, 272.

²⁸³ Denison, "Some Major Issues in Productivity Analysis: An Examination of Estimates by Jorgenson and Griliches."

calculations and their final result was that growth from the residual had been 1.03 per cent per year over 1950-62, as compared to Denison's figure of 1.37 per cent.²⁸⁴ There was no dispute about theory; all agreed that marginal products could be known from prices and that the inputs, or factors, should be limited to labour and capital, though there could be several types of either.

None of these procedures takes into account Pasinetti's point, that capital is a product as well as an input, so they neglect the effect of technical progress on its production. Abramovitz made much the same point in his review of Denison's estimates, where he questioned whether Denison was '... right in excluding the effect of quality change from his index of capital input ...'.²⁸⁵ Using numerical examples, Read showed that, because it neglects this point, Solow's residual is misleading; if, for example, the technical progress were to be confined to the production of capital goods, it would show no technical progress.²⁸⁶ Taking up Read's point, Rymes showed, what was implicit in Solow's article, that identifying technical progress with the residual after deducting the growth of output attributable to the increases of factors means that it depends on the factor shares.²⁸⁷ Hence, two economies in steady states with the same production functions, but not Cobb-Douglas, and the same rates of growth of the labour force, capital per head and output will still have different rates of technical progress if their wage and profit rates are different.

Rymes proposed that, just as the part of output per head attributable to factor increases is obtained by deducting the technical progress, the growth of capital per head should be adjusted by deducting the technical progress in the production of capital and he showed that if the two rates of technical progress are the same and constant, i.e. Harrod neutral technical progress in a steady state, that rate is independent of factor shares. In Solow's model output and capital are each treated as the same good, but Rymes demonstrated that his adjustment gives the same independence from factor shares whether there are one or several capital goods in the production of one consumption good, also when goods enter into their own production. Similar adjustments can be made to the estimates of Denison and Jorgenson and Griliches. That would make them more logical on the assumption of the same rates of technical progress for capital and final goods, but does not go so far as to treat capital goods as durable and heterogeneous.

Apart from the objections to the theory, there are objections to the statistical procedures of both Denison and of Jorgenson and Griliches that

²⁸⁴ Jorgenson and Griliches, "Issues in Growth Accounting: A Reply to Edward F. Denison," 89. Table 25.

²⁸⁵ Abramovitz, "Economic Growth in the United States," 772.

²⁸⁶ Read, "The Measure of Total Factor Productivity Appropriate to Wage-Price Guidelines."

²⁸⁷ Rymes, *On Concepts of Capital and Technical Change*.

are unrelated to their disagreements. What these procedures have in common is that they go from assuming that the marginal relations between prices, inputs and outputs that apply to individual goods and factors also apply to aggregates. Thus, it is assumed that dividing the value of a set of goods by a price index to obtain an index of quantity is the same as dividing the value of a single good by the price to give its quantity, and, the price being the marginal cost of producing that good, the price index can be presumed to be the marginal cost of increasing output as measured by the quantity index. This is something that would have to be demonstrated case by case. For example, if motor vehicles were to be aggregated, the value of the output divided by the price index should not be supposed to indicate the marginal cost of the increase of output of motor vehicles from one period to the next, for that cost would depend on the composition of the increase, whereas the price index would for practical reasons include vehicles from previous years with weights that may not correspond to recent sales. Such an error may be thought small, but it can accumulate over time even if it is. Moreover, the fewer the goods in any aggregation the more the composition is likely to change each period and the bigger the error to be expected. Hence, attempts to improve the estimates by having more aggregates with fewer components and more index numbers may increase the likelihood of error in the estimation of marginal costs. No attempt seems to have been made to estimate the error, presumably because it is considered small, but the residual being estimated is also small and must be shown to be bigger than the errors and not part of Abramovitz's "grand legatee".

R&D as a cause of technical progress

When their first estimates seemed to show that technical progress accounted for little of economic growth Jorgenson and Griliches were led to make a point that seems not to have been made before, namely that there had been progress in production techniques and the types of consumption goods available, but it was the consequence of investment in research and the technical development of goods (R&D) fetching a normal return, hence not the residual with which technical progress had been identified. They remarked that there were sceptics who preferred to call the residual a measure of ignorance and they concluded '... not that advances in knowledge are negligible, but that the accumulation of knowledge is governed by the same economic laws as any other process of capital accumulation.'²⁸⁸ Taken literally the statement makes three assertions relevant here. One is that the production of knowledge should be treated as a type of capital accumulation, the second that knowledge can be treated as though it were an ordinary good and the third that, since the residual is the

²⁸⁸ Jorgenson and Griliches, "The Explanation of Productivity Change," 274.

only measure of technical progress, part of what has normally counted as technical progress cannot be quantified. To the extent that knowledge is produced, does accumulate and is generated by R&D at a cost, the first assertion is valid. But the second is misleading. New knowledge is not lost and the usual assumption, with which Jorgenson and Griliches agree, is that, once generated, it is available to all, unless there are arrangements that can prevent that. Knowledge may resemble capital in being produced but is not a factor in the sense of Jorgenson and Griliches; reducing the quantities of factors reduces output, whereas the increase knowledge brings about of the quantity of output possible or of the ability to improve goods is permanent. Jorgenson and Griliches did not take up the questions raised by the third assertion, namely if technical progress results from the knowledge generated by R&D, how it is to be measured. Neither did Denison, who seemed not to accept technical progress as the outcome of investment in R&D.²⁸⁹ By the definition given above, technical progress can, in principle, be entirely the result of R&D and the residuals just described no more than the consequences of errors of theory and statistical procedure.

R&D is so obviously the main source of technical progress in reality that, for present purposes at least, other sources can be ignored. It is also obvious that much of it consists of improving existing goods and is, therefore, a source of competition. Firms that make goods that are improved by R&D compete to produce new, improved versions knowing that their competitors are doing the same and that, as new versions come out, older versions, being inferior, will have to be sold for less and will in time become unprofitable, by when their producers will have stopped them to free production capacity for newer versions. Virtually all such goods have numerous characteristics that users like or dislike and which vary from one version to another. For example, different types of bicycle are used for ordinary conveyance, for long tours, for racing on smooth surfaces and for rough terrain, and each firm making bicycles may make several versions of one or more types, each version with its own technical characteristics.

Now technical progress must be distinguished from the residual of Solow, Denison and Jorgenson and Griliches and from Arrow's "learning by doing" in that those who undertake R&D do so deliberately to generate knowledge useful to them at a cost. Its cost, being deliberately incurred, is an investment made by firms for profit, made possible by intellectual property rights, i.e. patents, copyright and trademarks, which gives them the means of preventing others from using knowledge obtained from their R&D. Every time a firm's R&D results in an improvement of some characteristic that firm

²⁸⁹ Denison, "Some Major Issues in Productivity Analysis: An Examination of Estimates by Jorgenson and Griliches," 62–63.

has a quasi-monopoly. It is not a full monopoly since its competitors produce the same good, but different versions with different combinations of characteristics, and each firm, as long as it keeps on producing new and better versions, can obtain a profit on its quasi-monopoly before the knowledge is superseded. But for patents, firms that have incurred the costs of R&D would be undercut by those that have not, for much knowledge generated by R&D cannot be kept secret for long. A good may reach a stage when the gains from R&D are too small to be worthwhile for producers. At this stage, as patents and intellectual property rights lapse with time, the good can be produced competitively by any firm. Or the costs of improvement may be so great that they can only be profitable if the firms producing the good are big enough. Since the improvements brought about by R&D are likely to be independent of the size of the firms making them, returns to scale are not constant when the cost of R&D is included.

R&D without durable, heterogeneous capital: endogenous growth models

If production is represented by functions of labour and undifferentiated, malleable capital, competition as described above among firms producing differentiated products must be confined to consumption goods. Then each version of a good has a demand curve, determined in part by the versions competitors are producing and by income distribution, and the firm producing it, as a quasi-monopolist with a given stock of capital, chooses the output that maximises profit. If the demand curve is known the firm has, in principle, a straightforward optimisation problem. To pay for the R&D the profit must exceed the marginal product of capital.

Rymes's method of adjusting the quantity of capital for technical progress in its production cannot be used, for there is no residual from the R&D, except with the third form of technical progress. If there is to be technical progress arising from R&D embodied in capital and the capital is to be malleable, the capital has to be differentiated in some way, which must be by supposing that R&D results in new types of capital and that there is no technical progress with the capital in existence. Romer devised a model like this.²⁹⁰ He assumes a single final good of which the part not consumed, which he terms capital, is transformed into producer durables in a fixed proportion. New types of these durables come into being through R&D carried out by trained workers (human capital). Only the newest durable is produced each period. Production of the final good is represented by a function of labour and one quantity, which is the sum of the quantities at the time of each type of durable raised to a power less than one, say zeta. As new durables are added that quantity increases and so does output. Depreciation can be allowed for but is left out for simplicity.

²⁹⁰ Romer, "Endogenous Technological Change."

Romer's distinctions between the consumption good and capital and between capital and producer durables are illusory, for there are no costs of going from the one to the other. He refers to capital as trucks, trains, computers and communication networks.²⁹¹ Elsewhere he refers to "blast furnaces, lathes, fork lift trucks, looms, etc."²⁹². But he does not explain the meaning of a fork lift truck raised to the power of zeta. There can only be a meaning if that object is a quantity of some substance. Then, would a better fork lift truck be more capital? Even if technical progress reduced its cost of production? And, since the object is added to a blast furnace raised to the same power, both objects must be of the same substance, a special form of malleable capital that can be cut into segments that, after having their quantities raised to the power zeta, can be added to give a new quantity of capital, which, in the model is produced by the same process as the consumption good.

Equally illusory is Romer's technical progress; his model is really one of accumulation. There is no technical progress in the durables, themselves; each type is like the other, except for the date it is added. As remarked by Aghion and Howitt, there is '... no obsolescence; new products are no better than existing ones.'²⁹³ The production function is independent of time and is symmetric with respect to the types of durables; if the quantities of different types are the same, the order in which they were added makes no difference, whereas the notion of technical progress implies that the later types should be the more productive. Zeta being less than one, the output of the final good is made greater if the output of each new durable is divided into smaller batches. Alternatively, instead of adding new types, one durable can be added repeatedly. The R&D is redundant.

Romer's technical progress is an illusion created by assuming that zeta is less than one and that the rate at which new durables are invented is proportional both to the numbers of trained workers in R&D and to the measure of the number of different durables already in existence. He contends that R&D becomes more productive with the accumulation of the knowledge it generates, which only means that the rate at which new durables come into being rises. Since the new durables are no different to the old ones, the rate at which they are invented makes a difference only because the later the period a given quantity of durables is produced, the more different types it is divided into and, zeta being less than one, the greater the addition to the production function.

This can be illustrated by taking time as discrete intervals, which is how Romer begins his exposition. From some point on, because of the rising rate

²⁹¹ Romer, 81.

²⁹² Romer, "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions," 31.

²⁹³ Aghion and Howitt, "A Model of Growth Through Creative Destruction," 392.

of new types of durables, several new ones are invented in a single period. Then, comparing two periods and assuming the quantity of the final good made into durables is the same in each and that within each period the amounts of the various types are equal, the later period adds a bigger quantity to the production function because the same quantity of the final good is divided into smaller packets and the packets are all raised to the power zeta. Were zeta one the two periods would add the same amount to the production function and were it zero the amounts added would be the number of packets, which is greater in the later period. Romer's use of a Cobb-Douglas production function adds to the oddity, for zeta represents the share of capital and that implies that the greatest technical progress occurs when the share of capital is zero and the least when it is the only factor. Changing to continuous time alters none of this.

Several models were devised after Romer's with the same intention, namely that of showing that, in free, competitive markets with rational consumers, investment in R&D is endogenously determined by firms maximising profits and, therefore, so are the technical progress arising from the knowledge it generates and the consequent economic growth. Rational behaviour in these endogenous growth models means that consumers have utility functions, which they maximise over time. Producers and consumers are assumed to know all they need to know about the future to make correct decisions, which, if R&D is a stochastic process, includes the extraneously given probabilities.

Endogenous growth models are formulated mathematically and must, therefore, be simple enough for the mathematics not to be intractable. One simplification is to confine the scope of technical progress to consumption goods, or to their production or to the process of producing the goods used in their production, but to only one of these forms in any model. Another is to leave out consumer durables; every consumption good is used once. Consumers are assumed to be alike; apart from having the same preferences and the same income, they must live equally long, best brought about by assuming they are immortal. Production, too, must be kept simple. Firstly, the prices of goods used in production must depend on labour costs and not include profit. Durable capital goods, unless malleable, are excluded because the profit rate enters into their prices. So, too, are intermediate goods, inputs that are used only once, if they are produced by means of labour and manufactured goods. Hence, when production is not represented as a function of malleable capital and labour, it is represented as a function of labour alone, or of labour using intermediate goods made by labour alone, or of intermediate goods without labour. In addition, when a good has several versions all are assumed to have the same factor and R&D costs. Secondly, if the technical progress occurs in production, there is only one

consumption or final good. This avoids having several different forms of capital or intermediate goods at the same time. No endogenous growth model does without these simplifications.

Another source of unmanageable mathematics is consumer choice of the kind described earlier, with several versions of the same good catering to different preferences and incomes. All endogenous growth models in which consumption goods are improved through R&D simplify with the assumption that different versions of a good can be ranked unambiguously as better or worse, allowing a single index as identifier. Then, when there are several versions of a good, each consumer chooses only one and, since consumers are alike, all choose the same one. Lancaster's proposal of attributing several characteristics to each good is ruled out, except when so restricted that goods can be ranked, an example being Stokey's model of learning by doing.²⁹⁴ In that model each good has associated desirable characteristics and each new version is the same as the previous one with the addition of one more characteristic.

Endogenous growth models and reality

This simplification leads to two unrealistic conclusions; first, that all goods are produced by monopolists and, second, that, if all firms are alike, no firm actually producing the latest version of a good invests in R&D. The first holds for consumption goods because of the restrictive assumptions that only one version is the latest and that consumers are alike, for the producer then maximises profit by pricing the latest version just low enough for previous ones to become obsolete, i.e. Bertrand competition resulting in a monopoly. It holds normally for capital and intermediate goods with the usual assumption of constant returns to scale, for producers are necessarily alike since they all maximise profit. Second, if all firms that carry out R&D are equally likely to devise the next version, regardless of whether they actually produce the good or not, the one that devised the version being made at the time does not invest in further improvement of that good while its monopoly lasts, for it would thereby shorten the period of monopoly and incur a cost. Alternatively, if it does develop what would be the latest version at the time and postpones production until its current monopoly ends, the R&D of its competitors in the meantime reduces the expected duration of that version. Each new version is, therefore, made by a different monopolist. Yet, in reality monopolies are rare and almost all new versions of existing goods are the outcome of the R&D of firms that have been producing those goods for some time and intend to continue their R&D to produce newer versions.

²⁹⁴ Stokey, "Learning by Doing and the Introduction of New Goods."

Segerstrom and Zolnierek remarked of the second conclusion that it is 'strongly counterfactual'.²⁹⁵ They devised a model in which firms are not alike; a firm that has been producing a good is assumed to be more likely, for given expenditure on R&D, to find the next version than other firms. In this case, in place of a set of consumer goods, the goods in question are the intermediate goods, which are used with labour to make the single final good according to a production function. Both the intermediate goods and R&D, which increases their productivity, consist solely of the final good. With a suitable choice of coefficients, Segerstrom and Zolnierek obtained the desired conclusion, that the firm that has been producing the latest version invests in R&D and is more likely than other firms to produce the next version. Nevertheless, eventually every firm producing a good is, by the laws of probability, replaced by a newcomer; instead of it happening immediately, there is a stochastically determined delay.

Choosing coefficients in this way to reach the result wanted may seem a minor modification of the mathematics of the model but, as a question of method, it is arbitrary and can be expected to create new problems. In this case a new problem is that it has an economic meaning; it implies a second type of knowledge, in-house knowledge of which each firm has its own, is acquired instantly at no cost, cannot be transferred and is extinguished when another firm produces the newest version. It is a departure from the principle common to all these models, that knowledge once generated is known to all. In-house knowledge does exist in reality and acquiring that of other firms is sometimes a reason for take-overs and mergers, but its characteristics are not as in the model. Moreover, it is rare in reality that a firm producing a good is replaced by a newcomer. Firms and industries do cease to exist as new goods come into existence, and firms close when they fail to make goods that are competitive with those of established competitors, but seldom are established firms replaced by new firms making new versions of the same goods. Finally, producing firms are still monopolies in the model.

If technical progress is the cause of increases of income or output per head and is faster the more knowledge R&D generates, it is to be expected that bigger economies with more R&D should grow faster than smaller ones and that growth should accelerate. Jones pointed out that such scale effects are what a number of endogenous growth models lead to and that they are contrary to what has been observed.²⁹⁶ Since 1950 the numbers of scientists and engineers in the high wage countries have grown faster than the economies and workforces of these countries, yet growth rates have not increased. Jones also concluded from his calculations of total factor

²⁹⁵ Segerstrom and Zolnierek, "The R&D Incentives of Industry Leaders," 745.

²⁹⁶ Jones, "R & D-Based Models of Economic Growth."

productivity in the US that the rate of increase of productivity did not seem to rise with time. It is also not true that bigger economies grow faster than smaller ones.

Jones's own explanation of why greater amounts of R&D had not led to faster growth was, in contrast to Romer, that, as knowledge accumulates, additions to it become harder, mainly because 'the most obvious ideas are discovered first', but also because of 'duplication and overlap of research'.²⁹⁷ His model is like that of Romer, with a single final good produced according to a function of labour and durables, each of which is made from the same amount of the final good.²⁹⁸ The durables are indexed according to the order of their invention and R&D, which is done solely by labour, adds new ones so that the index of the latest one is the measure of knowledge or technical progress. By choosing the coefficients relating the labour used for R&D to the change of knowledge suitably the rate of increase of productivity for a given amount of R&D can be made to decline as productivity increases and can be made to be less than proportional to the R&D.

But, as Jones pointed out, his model's growth in the steady state is that of the population. People optimise consumption over time and it takes ever more R&D to obtain the same increase of output, so they gradually reduce the amount of consumption they sacrifice to R&D. Like all proponents of endogenous growth models, Jones refers to steady states for his conclusions, so his model replaces the faster growth rates of bigger economies by stagnation.

Young proposed another explanation, namely that a bigger economy uses a wider range of goods for production and that its R&D, because it is spread over more goods, improves each good less.²⁹⁹ His model has a single final good produced by intermediate goods according to a symmetric function of these goods, their quantities multiplied by their productivity parameters. The intermediate goods make a continuum ranging to infinity, but only a finite part of that range is used, because there is a fixed cost to the use of each intermediate each period. Since the cost of improving productivity is assumed to increase more than proportionately to the improvement, production is spread over a range of intermediate goods. In the steady state the rate of growth of production is independent of the size of the economy or labour force. Young points out, however, that output per head is greater for a bigger economy, i.e. that the scale effect applies to the level and not to the rate.

²⁹⁷ Jones, 765.

²⁹⁸ Jones, Appendix.

²⁹⁹ Young, "Growth Without Scale Effects."

Howitt devised a model in which goods are improved and new ones are invented by modifying Young's model. He assumed that intermediate goods that have not already been used have to be invented, so that R&D is needed to add to the intermediate goods in use.³⁰⁰ His model has one final good, also produced according to a symmetric function of quantities of intermediate inputs multiplied by productivity parameters, and these inputs are, again, a continuum and are all made by the same amount of labour per unit. Both consumption and R&D consist of the final good, some of the R&D going to improving the productivity of the intermediate goods in use and the rest to inventing new ones. Howitt assumes that the improvements are stochastic and that the highest level of productivity increases in proportion to the ratio of R&D for productivity to the range of intermediate goods in use; the proportionate growth rate of this level of productivity is being assumed to be in inverse proportion to its level and in proportion to the R&D per unit of intermediate good in use. In the steady state in which the growth rate of the highest level of productivity, the ratio of labour to the range of intermediate goods in use and the proportion of the final good used as R&D to add new intermediate goods are all constant neither the growth rate nor consumption per head depends on the scale of the economy.

Quantifying knowledge and R&D

Each of these three models is an attempt to formulate an intuitively plausible idea as an endogenous growth model to obtain a specific desired result. But are these ideas well founded and what do the models add to them?

Jones's assertion that new knowledge becomes harder to obtain as knowledge accumulates seems reasonable on the analogy to the classical theory of economic rent from the extensive margin of cultivation of land; those parts of the land that yield the most for the effort are the first to be cultivated. But the analogy is false. When a patch of land is chosen for cultivation the choice is made with knowledge of the relation between yield and effort for all the different patches available. No such knowledge can be assumed of a relation between future research effort and its results. Some idea of what can be expected of research in the present or the near future is normal, for research is not undertaken blindly, especially since it has a cost. The right analogy would be to say that the yield of a known patch of land gives some expectation of the yield of the land next to it; the expectation would be reasonable, though it could be wrong and, barring other information, would be irrelevant to land further away.

Young and Howitt assumed that productivity increases with the amount of R&D and here the analogy is with the classical theory of economic rent

³⁰⁰ Howitt, "Steady Endogenous Growth with Population and R. & D. Inputs Growing."

from the intensive margin of cultivation; on any piece of land the yield increases with the effort with diminishing returns. This analogy is false for the same reason. Not only do the models assume that the relation through the future between R&D and the productivity is known for the various capital and intermediate good inputs into production in use, but it is also assumed that the relation is known for inputs that have not yet been used or have still to be invented.

In one respect the two analogies have a use; they draw attention to how the objections to the aggregation of capital apply to R&D in endogenous growth models. The term effort used above in describing the classical theory of rent implies that the various inputs into cultivation of land can be unambiguously represented as a single quantity, although they can consist of, among other things, labour, various kinds of machinery, fertilisers, pesticides and products of infrastructure like irrigation. Effort is, then, like capital, not the simple quantity it was taken to be by classical economists, such as West, Ricardo and J. S. Mill. In all endogenous growth models R&D is a single quantity, either an amount of labour or of the final good, although, as Jones, Romer and the other proponents of endogenous growth models have pointed out, the ability to do research has always depended on the products of earlier research. It is not simply that new knowledge is added to existing knowledge, what is termed “spillover” in these models. All scientific and engineering research depends on apparatuses and materials, which, themselves, were made possible by scientific research and which may be the main element of cost of any research. In other words, R&D is conducted by workers with various types of training using apparatuses and materials and cannot be represented as a single quantity independent of wages and profit for the same reason as capital.

The observation that firms compete in the production of goods with a variety of characteristics and invest in R&D to improve on them yields the evident explanation of why the numbers of scientists and engineers in the industrial countries have risen as proportions of the populations, which is what Jones wanted to explain, and reconciles those numbers with the estimates of investment in research in the US used by Howitt, which do not show such a tendency. Continual investment in R&D results in more types of complex goods and higher degrees of complexity, and the manufacture and maintenance, especially of the manufacturing equipment, accordingly require more suitably trained workers. Jones was right in drawing attention to the numbers, but wrong in associating them primarily with R&D. Howitt’s assumptions are contrived to give a steady level of R&D, but do not allow conclusions about reality and the degree to which the level of R&D has fluctuated in the US.

Endogenous growth models, interconnectedness and durability of capital goods

Production as represented in these models does not take account of the interconnectedness of the various parts of an industrial economy. In a developed industrial economy a good produced in one part of the economy or sector can enter, directly or indirectly, into the production of a good in another part or sector. Steel, for example, is part of practically all machinery and means of transport. Electric power, apart from being the energy source of many forms of transport, enters into the production of petrol and diesel engines, and all industrial products are transported at some stage. Advances in computational ability make for better design of machines and their components and better organisation of production and transport. So, directly or indirectly, steel, electric power and computers enter into the production of virtually all manufactures and can be called basic. Some goods enter into the production of some goods and not others and some, like most consumption goods, do not enter into production at all. Hence, technical progress in metallurgy, in the generation and transmission of power and in computers and computational techniques have influence throughout the economy, whereas a new soft drink only affects some consumers.

An implication is that the consequences of a technical advance with regard to a good depend on how the good enters into the production of other goods. An advance in steel metallurgy may result in cost reductions or quality improvements for many other goods, more so than an advance in the production of leather. A good may not be improved, but its production may be because of advances in what goes into its production. Representing production as a function of handmade intermediate goods does not allow sectors to be interconnected. Technical advances with respect to some intermediate goods have no effect on others and where R&D has not resulted in advances the goods remain backward. If one of these backward goods were to be chosen, as happens in endogenous growth models, it would have to become as advanced as the others instantly. In these models, to avoid having R&D taking a good through several periods of technical improvement, it is assumed that the technical level of all goods is automatically advanced to the most advanced level of other goods or somewhere near it when they are produced even if it is for the first time. Thus, in place of the pervasiveness of technical progress that occurs necessarily in an industrial economy there is an assumption, one that depends on another assumption, that the levels of technical advancement of different goods are quantified on the same scale, as though improvements of steel, leather and computer algorithms can be compared quantitatively. This is merely an elaboration of Pasinetti's point.

The alternative to representing production by a function of handmade intermediate goods is to represent it by functions of labour and malleable capital. Since technical progress does not change the substance of the capital, but is represented by multiplication of the labour and capital by a parameter, there is no need to discuss this alternative further.

Another consequence of representing production by handmade intermediate goods and no fixed capital is to obscure the difference between the results of R&D and the rate at which they are adopted through investment in productive capacity and infrastructure. Malleable capital, intermediate goods and labour adapt immediately to the new production and a new firm that starts producing a new or improved good produces instantly. No time and money are spent on creating or adapting productive capacity and firms only risk their investments in R&D, unlike reality, in which the cost of the capacity to produce a new or improved good may be several times the cost of the R&D and the main determinant of whether or not the good is made and when. If fixed capital is assumed away, old goods are replaced by new ones whenever the cost of the R&D can be profitably regained. Old equipment does not exist alongside new equipment; a firm's plant older than a competitor's is scrapped and old bridges are instantly replaced when better new ones are profitable.

The purpose of endogenous growth models

Perhaps some proponents of endogenous growth models, rather than dispute the criticisms, would put the purpose of their models as showing how, if markets are free and consumers rational, firms will invest in R&D and bring about economic growth. If that is the purpose, it is inconsistent with the efforts that have been made to devise models that conform to some of the observed facts and is at the same time a justification of attempts to demonstrate something that is not true. The state has long been responsible for a large part of the R&D in every industrial economy and many technical advances, like radar, the jet engine, the computer and many medicines, have depended on research by state institutions or by institutions, including universities and private firms, working on behalf of the state.³⁰¹ It is unnecessary to argue this in detail, for it is one example of the point being made here, that technical progress cannot be understood or satisfactorily described if production is not represented as labour using durable, heterogeneous capital goods.

Still, the academic question remains, is it possible to achieve that purpose by adding R&D to a model of intertemporal general equilibrium like that expounded by Arrow and Hahn? If it is, it would avoid the objections

³⁰¹ Mazzucato, *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*.

to the kinds of models discussed here in the same way as the model of Arrow and Hahn avoids the objections to production functions, equally implausibly. But there is no reason to believe that R&D is compatible with the assumptions about production used in the proof that there is an equilibrium. R&D is like capital in that its cost is composed of wage and profit, but its effects on production are different and there is no assurance that they can be accommodated in the assumptions. Investment in capital goods allows more production, whereas R&D is the use of labour and goods to alter the goods available and the processes for making them, and does this irreversibly. In addition, if there are no laws on intellectual property the benefit from the R&D does not accrue solely to whoever incurred the cost, and if there are the production possibilities are not the same for all. The durations of patents and copyright are determined by law and can change and do vary according to product. As pointed out earlier, because the representation of production in such models is a mathematical abstraction, it is not always possible to say if a particular form of production is compatible with the assumptions. For instance, there is no formula that tells to what extent departures from constant returns to scale and externalities are compatible with the model. The same is true of the compatibility of R&D. Perhaps mathematical research will one day supply the knowledge lacking, and it may show that the R&D compatible with the existence proof of general equilibrium is too small to be used for discussing economic growth. As a matter of economic theory, investment in R&D is obviously related to investment in productive capacity and is not determined by calculations of future profit with perfect foresight.

Finally, unrelated to the question of how production is represented, there is the oft made point of the illogicality of models of a distant future determined by technical progress, which is, by its nature, not predictable, except roughly for the near future. Proponents of endogenous growth models like to draw their conclusions from steady states, presumably in some indefinite future, and, therefore, simplify their depictions of technical progress to yield steady states. To be able to do this they confine technical progress to improvements and inventions that differ from those preceding them only by a factor, if at all, a sameness that is the antithesis of technical progress. They go so far as to call the replacement of old versions of goods by improved versions 'Schumpeterian creative destruction'. Schumpeter, who insisted on the unpredictability of inventions and their disruption of the equilibrium into which the economy moved between inventions, would have objected. For him the steady state was stagnation and its disruption the occasion for entrepreneurship, profit and growth.

3. SUMMING UP

At issue is whether or not representing economic production as functions of quantities of capital can be used to describe or explain the functioning of real economies, and the answer is that any assertion about economic reality that depends on functions of quantities of capital is wrong. Three reasons have been given. One is that of Sraffa, that the value of the stock of capital is not a quantity independent of wages and profits but is determined by them. A second reason derives from technical progress and was first referred to by Pasinetti. The third derives from international trade. They are logically distinct, though the second and the third only became apparent because of the first.

It follows from the first reason that production relations alone do not determine wage and profit rates. This indeterminacy can be removed by introducing demand. For instance, wage and profit rates in steady states are determined by the saving and consumption rates. One degree of indeterminacy may remain, for employment is not necessarily determined. Some of the marginal relations derived from production functions can hold for comparisons between steady states, which means they are mathematical relations, not relations of cause and effect. And they cannot be used to describe change over time outside steady states, for wages and profits change and cause the plant and machinery that are fixed capital and relative prices to change, too. Such changes are too complicated to describe or for households and firms to calculate optimal expenditure and production, and, as is evident, real economies are not in equilibrium, for expectations are not all realised. Intertemporal general equilibrium models can show, with some assumptions that seem general and others impossible, that there is equilibrium with full employment apart from steady states, but these models are just mathematical existence theorems unrelated to reality. Their very generality and abstractness obscure the extent to which the assumptions can allow for the complications of reality. It is not known, for instance, to what extent their assumptions can accommodate R&D and the technical progress it results in. These models are also illogical in the sense that they require perfect foresight, though technical progress is inherently unpredictable, certainly over the long run.

There are also the models and estimates of technical progress using aggregates which equate technical progress to the part of increased output or GDP not attributable to increased use of factors. There is, as yet, no objective way of quantifying technical progress in the form of improvements of consumption and capital goods and the invention of new ones, nor the R&D of which it is the result. Notional adjustments of output and GDP can be made and may be necessary, but they are not objective. Hence, these models

and estimates either omit some technical progress or, if the notional adjustments are made, rely on supposition. Technical progress defined like this, whether adjusted for technical progress in the production of capital or not, is only what the models cannot explain, besides which there is no independent check that its estimates are not merely the outcome of the errors and approximations of the data and their treatment. R&D should be allowed for in any residual by deducting its cost, which can leave little or nothing, and whatever may be left can still be technical progress from the same source. Then there are the endogenous growth models, which do not provide means for estimating technical progress but have as their purpose to show that firms in competitive free markets will invest in R&D and reach a steady state. They need the assumptions that R&D can be quantified, that goods are made by hand or are same as the final good and that firms and consumers foresee the future goods, which are just the present goods multiplied by a known parameter, sometimes unity.

Since most international trade is in manufactures, it is to be expected that technical progress should be one of the main determinants of the pattern of trade. Most tradable manufactures are goods made by firms competing among themselves by using R&D to design improved versions of the goods they make. The competition is international and countries export versions made by their firms and import versions made abroad. Firms are prevented from copying the versions of competitors by the laws on intellectual property, besides which much of the technical knowledge of firms is in-house and not available to others. A firm starting to produce a good of this kind will not have technical knowledge equivalent to the patented and in-house knowledge of established producers, which continue their R&D and keep their lead. Low wage country firms that do not have up to date technical knowledge may still be able to compete by producing cheap versions inferior to those of high wage country firms. But, to reach the level of these firms, they must give the established producers enough inducement for them to impart their knowledge despite their aversion to creating new competitors, for instance by allowing access to promising markets. If they cannot do that, low wage country firms are confined to competing among themselves with the same simple goods and, perhaps, cheap versions of goods produced in high wage countries, i.e. low wage goods with little value added.

Much of manufacture is accounted for by firms with proprietary specialised knowledge, be they big multinational firms or small, and they persist because of their R&D, contrary to the endogenous growth models. A function of brand names is to signal this. In addition, much R&D and scientific research are carried out by state and other organisations. One consequence is that the firms, especially the biggest, that constitute much of the industry of a country came into existence under a variety of

circumstances, not because of factor endowments. A firm that began with one good may have changed out of recognition through diversification, acquisitions, mergers and changes of its original products. Another consequence is that the relation between the degree of advancement of an economy and its competitiveness is loose, for competing firms almost certainly do similar R&D. In rough terms, the more different versions of a particular good there are, the greater the amount of R&D. It is also possible for the firms of two countries producing the same good to be equally advanced, but for those of the one to produce more versions than those of the other. In that case the first country is likely to be more competitive in the sense that it will sell more of that good, other things being equal. Such assertions are necessarily imprecise because they concern phenomena, notably technical progress and R&D, that cannot be quantified and that make dependable mathematical formulation of economies impossible, at least at present. This may seem unfortunate, but such assertions are not the less useful for being imprecise. Certainly more useful than models that provide formulae for equilibria in some indefinite future when it is hard to say what the state of an economy is in the present, models that seem just to be ends in themselves.

Far from being about theoretical abstractions, the disputes about capital were about the basic and practical questions of understanding how economies function and formulating policies. Ignoring the conclusions to be drawn from the disputes as mere abstractions on the grounds of being practical or pragmatic has had the perverse effect, that it is the models that have resulted that are abstractions remote from reality. Being practical and pragmatic is also often given as justification for simplifications that make the mathematics of models easier and more manageable, one aspect being that the results are those desired or intended, and the consequence has been that it has become normal to devise mutually irreconcilable models purporting to explain different aspects of the same phenomena and ignore the inconsistencies. Yet it is sometimes asserted that mathematically formulated models are rigorous and verbal arguments are impressionistic. The mathematics may be rigorous, but the economics is not. Things not quantifiable or too complicated to be amenable to mathematics can be discussed verbally with rigour; assertions need not be quantifiable to be precise and statements that cannot be proved can be made explicitly as conjecture. Assertions drawn solely from simplifications made to formulate such things mathematically are tautologies of the simplifications, conjecture at best, otherwise unrelated to things that are.

REFERENCES

- Abramovitz, Moses. "Economic Growth in the United States." *The American Economic Review* 52, no. 4 (1962): 762–82.
- Aghion, Philippe, and Peter Howitt. "A Model of Growth Through Creative Destruction." *Econometrica* 60, no. 2 (1992): 323–51. <https://doi.org/10.2307/2951599>.
- Ahmad, Ehtisham. "Why Is It so Difficult to Implement a GST in Pakistan?" *The Lahore Journal of Economics* 15, no. Lahore Journal of Economics. Special Edition. Sixth Annual Conference on Management of the Pakistan Economy (2010): 139–69.
- Alexander, Sidney S. "Effects of a Devaluation on a Trade Balance." *Staff Papers - International Monetary Fund* 2, no. 2 (1952): 263–78.
- Amsden, Alice H. *Asia's Next Giant*. Oxford University Press: Oxford University Press., 1989.
- Arrow, Kenneth Joseph, and Frank Horace Hahn. *General competitive analysis*. Amsterdam: North-Holland, 1983.
- Atkinson, A. B. "Import Strategy and Growth under Conditions of Stagnant Export Earnings." *Oxford Economic Papers* 21, no. 3 (1969): 325–38.
- Bacha, Edmar, and Lance Taylor. "Foreign Exchange Shadow Prices: A Critical Review of Current Theories." *The Quarterly Journal of Economics* 85, no. 2 (1971): 197–224. <https://doi.org/10.2307/1880701>.
- Balassa, Bela. "Tariff Protection in Industrial Countries: An Evaluation." *Journal of Political Economy* 73, no. 6 (December 1, 1965): 573–94. <https://doi.org/10.2307/1829884>.
- Balassa, Bela, and Daniel M. Schydrowsky. "Domestic Resource Costs and Effective Protection Once Again." *Journal of Political Economy* 80, no. 1 (January 1, 1972): 63–69. <https://doi.org/10.2307/1830131>.
- Bello, Walden F., and Stephanie Rosenfeld. *Dragons in Distress: Asia's Miracle Economies in Crisis*. San Francisco, Calif.: Food First, 1992.
- Brown, E. H. Phelps. "The Meaning of the Fitted Cobb-Douglas Function." *The Quarterly Journal of Economics* 71, no. 4 (1957): 546–60. <https://doi.org/10.2307/1885710>.

REFERENCES

- Bruno, Michael. "Domestic Resource Costs and Effective Protection: Clarification and Synthesis." *Journal of Political Economy* 80, no. 1 (1972): 16–33.
- Burmeister, Edwin. "Wicksell Effects." In *The New Palgrave: Capital Theory*. London: Macmillan, 1990.
- Castley, Robert. *Korea's Economic Miracle: The Crucial Role of Japan*. Palgrave Macmillan, 2014.
- Chamberlin, Edward. *The Theory of Monopolistic Competition*. Cambridge: Harvard Univ. Pr., 1933.
- Champernowne, D. G. "The Production Function and the Theory of Capital: A comment." *Review of Economic Studies* 21, no. 2 (54 1953): 112–35.
- Chang, Ha-Joon. *Kicking Away the Ladder: Development Strategy in Historical Perspective*, 2007.
- Chaudhry, Shahid Amjad. "Pakistan: Indus Basin Water Strategy – Past, Present and Future.," Vol. 15. Lahore: Lahore School of Economics, 2010.
- Chipman, John S. "A Survey of the Theory of International Trade: Part 1, The Classical Theory." *Econometrica* 33, no. 3 (1965): 477–519. <https://doi.org/10.2307/1911748>.
- . "A Survey of the Theory of International Trade: Part 3, The Modern Theory." *Econometrica* 34, no. 1 (1966): 18–76. <https://doi.org/10.2307/1909855>.
- Clower, Robert, and Richard Lipsey. "The Present State of International Liquidity Theory." *The American Economic Review* 58, no. 2 (1968): 586–95.
- Cohen, Avi J., and G. C. Harcourt. "Retrospectives: Whatever Happened to the Cambridge Capital Theory Controversies?" *The Journal of Economic Perspectives* 17, no. 1 (2003): 199–214.
- Curcuru, Stephanie E., Tomas Dvorak, and Francis E. Warnock. "Cross-Border Returns Differentials." *The Quarterly Journal of Economics* 123, no. 4 (2008): 1495–1530.
- Cyhn, Jin W. *Technology Transfer and International Production: The Development of the Electronics Industry in Korea*. Cheltenham [etc.]: Edward Elgar, 2002.

- De Cecco, Marcello. *Money and Empire: The International Gold Standard, 1890-1914*. Oxford: B. Blackwell, 1974.
- De Grauwe, Paul. *International Money: Postwar Trends and Theories*. Oxford: Oxford University Press, 1996.
- Denison, Edward Fulton. "Some Major Issues in Productivity Analysis: An Examination of Estimates by Jorgenson and Griliches." *Survey of Current Business* 52, no. 5. Part II (May 1972): 37–63.
- Dobb, Maurice. *Soviet Economic Development Since 1917. 6th Ed.* London: Routledge & Kegan Paul, 1966.
- Doner, Gregory W. Noble, John Ravenhill, and World Bank. "Industrial Competitiveness of The Auto Parts Industries In Four Large Asian Countries: The Role of Government Policy In A Challenging International Environment," 2007. <http://elibrary.worldbank.org/content/workingpaper/10.1596/1813-9450-4106>.
- Dooley, Michael, and Peter Garber. 'Is It 1958 or 1968? Three Notes on the Longevity of the Revived Bretton Woods System'. *Brookings Papers on Economic Activity*. 2005.
- Dore, Ronald Philip. *Structural adjustment in Japan : 1970-82*. Geneva: Internat. Labour Office, 1986.
- Dornbusch, Rudi., and F. Leslie C. H. Helmers. *The Open Economy: Tools for Policymakers in Developing Countries*. EDI Series in Economic Development. The World Bank. Washington, DC: Economic Development Institute of the World Bank., 1986.
- Dudler, Hermann-Josef. "Book Review: Balance-of-Payments Deficits and the International Market for Liquidity. *Essays in International Finance*, No. 46." *finarch FinanzArchiv / Public Finance Analysis* 25, no. 3 (1966): 541–43.
- Eichengreen. *Golden Fetters*. Oxford University Press : Oxford University Press.
- Ellis, Howard Sylvester., and Robert Miller Neal. *Readings in the Theory of International Trade: Selected by a Committee of the American Economic Association; Compiled by Howa., 1950*.
- Emmanuel, Arghiri. "A Note on 'Trade Pattern Reversals.'" *Journal of International Economics* 8, no. 1 (February 1978): 143–45. [https://doi.org/10.1016/0022-1996\(78\)90047-8](https://doi.org/10.1016/0022-1996(78)90047-8).

REFERENCES

- Emmanuel, Arghiri. *L'échange inégal : essai sur les antagonismes dans les rapports économiques internationaux*. Paris: F. Maspero, 1979.
- Enos, John Lawrence, and U-hui Park. *The Adoption and Diffusion of Imported Technology. The Case of Korea*. London: C. Helm, 1988.
- Ethier, Wilfred. "The Theorems of International Trade in Time-Phased Economies." *Journal of International Economics* 9, no. 2 (May 1, 1979): 225–38. [https://doi.org/10.1016/0022-1996\(79\)90005-9](https://doi.org/10.1016/0022-1996(79)90005-9).
- Felipe, Jesus, and Franklin M. Fisher. "Aggregation in Production Functions: What Applied Economists Should Know." *Metroeconomica* 54, no. 2 & 3 (2003): 208–62.
- Fischer, Stanley. "Devaluation and Inflation." In *The Open Economy: Tools for Policymakers in Developing Countries*, n.d.
- Fleming, Marcus J. "Reserve Creation and Real Reserves." International Monetary Fund, n.d.
- Friedman, Milton. "The Case for Flexible Exchange Rates." In *Essays in Positive Economics*, n.d.
- Gardner, R.N. *Sterling Dollar Diplomacy: Anglo-American Collaboration in the Reconstruction of Multilateral Trade*. Clarendon Press, 1956. <https://books.google.com/books?id=xksn2xq3FOoC>.
- Garegnani, P. "Heterogeneous Capital, the Production Function and the Theory of Distribution." *The Review of Economic Studies* 37, no. 3 (1970): 407–36. <https://doi.org/10.2307/2296729>.
- Giscard d'Estaing, Valéry. "La Politique Monétaire Internationale de la France." Institut d'Études Bancaires et Financières., 1965.
- Goldberg, Pinelopi Koujianou, and Frank Verboven. "The Evolution of Price Dispersion in the European Car Market." *ROES Review of Economic Studies* 68, no. 4 (2001): 811–48.
- Gourinchas, Pierre-Olivier, and Hélène Rey. *From World Banker to World Venture Capitalist: US External Adjustment and the Exorbitant Privilege*. NBER Working Paper Series, no. w11563. Cambridge, Mass: National Bureau of Economic Research, 2005. <http://www.nber.org/papers/w11563>.
- Government of Pakistan, Finance Division, Economic Adviser's Wing. *Pakistan Economic Survey. 1980-81*, 1982.
- . *Pakistan Economic Survey. 1999-2000*, 2001.

———. *Pakistan Economic Survey 2018-19*, 2020.

Government of Pakistan, Ministry of Federal Education and Professional Training, United Nations Education, Scientific and Cultural Organization, and United Nations Children's Fund. *Pakistan Education Statistics, 2017-17*. Islamabad: Academy of Educational Planning and Management, Ministry of Education, 2009.

Gresser, Edward. "America's Hidden Tax on the Poor. The Case for Reforming U.S. Tariff Policy." Progressive Policy Institute. Policy Report., March 2002.

Haberler, Gottfried von. *The Theory of International Trade: With Its Applications to Commercial Policy*. London: Hodge & Co., 1936.

Habib, Maurizio Michael. "Excess Returns on Net Foreign Assets: The Exorbitant Privilege from a Global Perspective." European Central Bank Working Paper 1158, February 2010. <https://www.ecb.europa.eu/home/search/html/index.en.html?q=+1158>.

Hahn, Frank. "The Neo-Ricardians." *Cambridge Journal of Economics* 6, no. 4 (1982): 353–74.

Haq, Mahbub ul. *The Strategy of Economic Planning: A Case Study of Pakistan*. Karachi: Oxford University Press, 1966.

Haque, Irfan. "The Rise of Bilateralism in Trade and Its Implications for Pakistan." *Lahore Journal of Economics* 14, no. Special Edition, September, 2009 (2009).

Harberger, Arnold C. "Currency Depreciation, Income, and the Balance of Trade." *Journal of Political Economy* 58, no. 1 (1950): 47–60.

Harrod, Roy Forbes. *International Economics*. London: Nisbet, 1933.

Helfferrich, Karl. *Das Geld*. Leipzig: Hirschfeld, 1923.

Helpman, Elhanan., and Paul R. Krugman. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition and the International Economy*. Cambridge (Massachusetts) [etc.]: MIT, 1993.

Howitt, Peter. "Steady Endogenous Growth with Population and R. & D. Inputs Growing." *Journal of Political Economy* 107, no. 4 (1999): 715–30. <https://doi.org/10.1086/250076>.

REFERENCES

- International Monetary Fund. *International Reserves: Needs and Availability; Papers and Proceedings, Seminar at the International Monetary Fund, June 1-3, 1970*. Washington, D.C., 1970.
- International Monetary Fund. "World Economic Outlook Database April 2011," n.d. <https://www.imf.org/external/pubs/ft/weo/2011/01/weodata/index.aspx>.
- Isard, Peter. "How Far Can We Push the 'Law of One Price'?" *The American Economic Review* 67, no. 5 (1977): 942–48.
- Johnson, Chalmers A. *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975*. Stanford University Press, 1982.
- Johnson, Harry G. "The Welfare Costs of Exchange-Rate Stabilization." *Journal of Political Economy* 74, no. 5 (October 1, 1966): 512–18. <https://doi.org/10.2307/1829598>.
- Jones, Charles I. "R & D-Based Models of Economic Growth." *Journal of Political Economy* 103, no. 4 (1995): 759–84.
- Jorgenson, Dale W., and Zvi Griliches. "Issues in Growth Accounting: A Reply to Edward F. Denison." *Survey of Current Business* 52, no. 5. Part II (May 1972): 65–94.
- . "The Explanation of Productivity Change." *The Review of Economic Studies* 34, no. 3 (1967): 249–83. <https://doi.org/10.2307/2296675>.
- Kemp, Murray C. "Heterogeneous Capital Goods and Long-Run Stolper-Samuelson Theorems*." *Australian Economic Papers* 12, no. 21 (December 1, 1973): 253–60. <https://doi.org/10.1111/j.1467-8454.1973.tb00309.x>.
- Kenen, Peter B. "International Liquidity and the Balance of Payments of a Reserve-Currency Country." *The Quarterly Journal of Economics* 74, no. 4 (1960): 572–86. <https://doi.org/10.2307/1884352>.
- Keynes, John Maynard. *The General Theory of Employment, Interest and Money*. Macmillan, 1936.
- Keynes, John Maynard. "The German Transfer Problem." *The Economic Journal* 39, no. 153 (1929): 1–7. <https://doi.org/10.2307/2224211>.
- Keynes, John Maynard, and Bertil Ohlin. "The Reparation Problem: A Discussion." *The Economic Journal* 39, no. 154 (1929): 172–82. <https://doi.org/10.2307/2224537>.

- Khan, Azizur Rahman. "What Has Been Happening to Real Wages in Pakistan?" *The Pakistan Development Review* 7, no. 3 (1967): 317–47.
- Kim, Lin-Su. "The Dynamics of Technology Development: Lessons from the Korean Economy." In *Competitiveness, FDI and Technological Activity in East Asia.*, edited by Sanjaya Lall, Shujiro Urata, and Yang Yao, n.d.
- Kindleberger, Charles P. *Europe and the Dollar*. Cambridge, Mass.: MIT Pr., 1969.
- . *International Money: A Collection of Essays*. London: George Allen & Unwin, 1981.
- Knetter, Michael M. "International Comparisons of Pricing-to-Market Behavior." *The American Economic Review* 83, no. 3 (1993): 473–86.
- Komiya, Ryutaro., Masahiro. Okuno, and K. Suzumura. *Industrial Policy of Japan*. Tokyo [etc.: Academic Press, 1988.
- Krause, Lawrence, and Sueo Sekiguchi. "Japan and the World Economy," 1976.
- Krueger, Anne O. "Evaluating Restrictionist Trade Regimes: Theory and Measurement." *Journal of Political Economy* 80, no. 1 (January 1, 1972): 48–62. <https://doi.org/10.2307/1830130>.
- Krugman, Paul R. "Equilibrium Exchange Rates." In *International Policy Coordination and Exchange Rate Fluctuations*, n.d.
- . "Pricing to Market When the Exchange Rate Changes." National Bureau of Economic Research, 1986. <http://www.nber.org/papers/w1926>.
- Laïdi, Zaki. *Enquete sur la Banque mondiale*. Paris: Fayard ;, 1989.
- Lancaster, Kelvin J. "A New Approach to Consumer Theory." *Journal of Political Economy* 74, no. 2 (1966): 132–57.
- Leamer, Edward E. *Sources of International Comparative Advantage: Theory and Evidence*. Cambridge, Mass. [etc.]: The MIT Press, 1987.
- Leontief, Wassily. "Domestic Production and Foreign Trade; The American Capital Position Re-Examined." *Proceedings of the American Philosophical Society* 97, no. 4 (1953): 332–49.
- . "Introduction to a Theory of the Internal Structure of Functional Relationships." *Econometrica* 15, no. 4 (1947): 361–73. <https://doi.org/10.2307/1905335>.

REFERENCES

- Levhari, David. "A Nonsubstitution Theorem and Switching of Techniques." *The Quarterly Journal of Economics* 79, no. 1 (1965): 98–105. <https://doi.org/10.2307/1880514>.
- Lewis, Stephen R., and Stephen E. Guisinger. *Measuring Protection in a Developing Country: The Case of Pakistan*. Cambridge, Mass.; Williamstown, Mass.: Project for Quantitative Research in Economic Development, Center for International Affairs, Harvard University; Center for Development Economics, Williams College, 1966.
- Lin, Ching-Yuan. *Industrialization in Taiwan: 1946 - 1972; Trade and Import-Substitution Policies for Developing Countries*. New York: Praeger, 1979.
- List, F. *Das Nationale System Der Politischen Oekonomie*. Edited by K.T. von Eheberg. Vol. 3. Friedrich List's Gesammelte Schriften. Cotta, 1851. <https://books.google.com/books?id=sWtnAAAACAAJ>.
- Lynd, Douglas. "The Education System in Pakistan: Assessment of the National Education Census." United Nations Education, Scientific and Cultural Organization, 2007.
- Mainwaring, L. "Relative Prices and 'Factor Price' Equalisation in a Heterogeneous Capital Goods Model." *Aust Econ Papers Australian Economic Papers* 15, no. 26 (1976): 109–18.
- Mazzucato, Mariana. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. London: Anthem Press, 2013.
- McKinnon, Ronald I. *Private and Official International Money: The Case for the Dollar*. Princeton, N.J.: International Finance Section, Dept. of Economics, Princeton University, 1969.
- Meade, James Edward. *The Theory of International Economic Policy. Vol. 1, The Balance of Payments*. London [etc.]: Oxford University Press, 1960.
- . *The Theory of International Economic Policy. Vol. 2, Trade and Welfare*. London u.a.: Oxford Univ. Pr., 1961.
- Metcalf, J.S., and Ian Steedman. "Reswitching and Primary Input Use." *The Economic Journal* 82, no. 325 (March 1, 1972): 140–57. <https://doi.org/10.2307/2230211>.
- Nasim, Anjum. "Determinants of Inflation in Pakistan." State Bank of Pakistan, 1996.

- Ohlin, Bertil. *Interregional and International Trade*. Cambridge, Mass: Harvard Univ. Press, 1957.
- Ozawa, Terutomo. *Japan's Technological Challenge to the West, 1950-1974: Motivation and Accomplishment*. Cambridge (Mass.): Mit, 1974.
- Papanek, Gustav F. *Pakistan's Development: Social Goals and Private Incentives*. Karachi: Oxford Univ. Pr., 1970.
- Pasha, Hafiz A., and Aisha Ghaus-Pasha. "The Future Path of Tax Reforms in Pakistan." In *Pakistan: Moving the Economy Forward*, 171–97, 2015.
- Pasinetti, Luigi L. "On Concepts and Measures of Changes in Productivity." *The Review of Economics and Statistics* 41, no. 3 (1959): 270–86. <https://doi.org/10.2307/1927453>.
- . "Switches of Technique and the 'Rate of Return' in Capital Theory." *The Economic Journal* 79, no. 315 (1969): 508–31. <https://doi.org/10.2307/2230379>.
- Pigou, A. C. *Protective and Preferential Import Duties*. London, 1936.
- Polak, Jacques Jacobus. *Money-National and International*. Paris: OECD, 1970.
- Raj, K. N., and A. K. Sen. "Alternative Patterns of Growth under Conditions of Stagnant Export Earnings." *Oxford Economic Papers* 13, no. 1 (1961): 43–52.
- Read, L. M. "The Measure of Total Factor Productivity Appropriate to Wage-Price Guidelines." *The Canadian Journal of Economics / Revue Canadienne d'Economie* 1, no. 2 (1968): 349–58. <https://doi.org/10.2307/133503>.
- Rist, Charles. *Essais sur Quelques Problemes Economiques et Monetaires*. Paris: Recueil Sirey, 1933.
- Robinson, Joan. "Banking Policy and the Exchanges." *The Review of Economic Studies* 3, no. 3 (1936): 226–29. <https://doi.org/10.2307/2967630>.
- . *The Economics of Imperfect Competition*. London: Macmillan, 1933.
- . "The Production Function and the Theory of Capital." *The Review of Economic Studies* 21, no. 2 (1953): 81–106. <https://doi.org/10.2307/2296002>.

REFERENCES

- . "The Production Function and the Theory of Capital – a Reply." *The Review of Economic Studies* 23, no. 3 (56 1955): 247.
- Rodrik, Dani. *What Does the Political Economy Literature on Trade Policy (Not) Tell Us That We Ought to Know?* Cambridge, Mass.: National Bureau of Economic Research, 1994.
- Romer, Paul M. "Endogenous Technological Change." *Journal of Political Economy* 98, no. 5 (1990): 571–102.
- . "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions." *Journal of Development Economics* 43, no. 1 (February 1, 1994): 5–38. [https://doi.org/10.1016/0304-3878\(94\)90021-3](https://doi.org/10.1016/0304-3878(94)90021-3).
- Routh, Guy. *Occupation and Pay in Great Britain 1906-1960*. Cambridge: Cambridge University Press, 1965.
- Rueff, Jacques. "Mr. Keynes' Views on the Transfer Problem." *The Economic Journal* 39, no. 155 (1929): 388–408. <https://doi.org/10.2307/2224179>.
- Rymes, Thomas K. *On Concepts of Capital and Technical Change*. Cambridge: University Press, 1980.
- Samuelson, Paul A. "Free Trade's Intertemporal Pareto-Optimality." *Journal of International Economics* 8, no. 1 (February 1, 1978): 147–49. [https://doi.org/10.1016/0022-1996\(78\)90048-X](https://doi.org/10.1016/0022-1996(78)90048-X).
- . "International Trade and the Equalisation of Factor Prices." *The Economic Journal* 58, no. 230 (June 1948): 163. <https://doi.org/10.2307/2225933>.
- . "Parable and Realism in Capital Theory: The Surrogate Production Function." *The Review of Economic Studies* 29, no. 3 (June 1962): 193–206.
- . "Prices of Factors and Good in General Equilibrium." *The Review of Economic Studies* 21, no. 1 (1953): 1–20. <https://doi.org/10.2307/2296256>.
- . "Trade Pattern Reversals in Time-Phased Ricardian Systems and Intertemporal Efficiency." *Journal of International Economics* 5, no. 4 (November 1975): 309–63. [https://doi.org/10.1016/0022-1996\(75\)90037-9](https://doi.org/10.1016/0022-1996(75)90037-9).
- Sargent, Thomas J. "The Ends of Four Big Inflations." In *Rational Expectations and Inflation*, n.d.

- Schwartz, Jacob T. *Lectures on the Mathematical Method in Analytical Economics*. New York: Gordon and Breach, 1961.
- Segerstrom, Paul S., and James M. Zolnierok. "The R&D Incentives of Industry Leaders." *International Economic Review* 40, no. 3 (1999): 745–66.
- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations: Vol. 1-2*. Edited by Edwin Cannan. London: Methuen, 1961.
- Soligo, Ronald, and Joseph J. Stern. "Tariff Protection, Import Substitution and Investment Efficiency." *The Pakistan Development Review* 5, no. 2 (1965): 249–70.
- Solomon, Robert. *The International Monetary System, 1945-1981*. New York [etc.]: Harper & Row, 1982.
- Solow, Robert M. *Capital Theory and the Rate of Return*. Professor F. de Vries Lectures. Amsterdam: North-Holland Publishing Company, 1963.
- . "Substitution and Fixed Proportions in the Theory of Capital." *The Review of Economic Studies* 29, no. 3 (June 1962): 207–18.
- . "Technical Change and the Aggregate Production Function." *The Review of Economics and Statistics* 39, no. 3 (1957): 312–20. <https://doi.org/10.2307/1926047>.
- . "Technical Progress, Capital Formation, and Economic Growth." *The American Economic Review* 52, no. 2 (1962): 76–86.
- . "The Production Function and the Theory of Capital." *The Review of Economic Studies* 23, no. 2 (1955): 101–8. <https://doi.org/10.2307/2296293>.
- Sraffa, Piero. *Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory*. Cambridge, 1960.
- Sraffa, Piero. "The Laws of Returns under Competitive Conditions." *The Economic Journal* 36, no. 144 (1926): 535–50. <https://doi.org/10.2307/2959866>.
- Steedman, Ian, and J. S. Metcalfe. "Reswitching, Primary Inputs and the Heckscher-Ohlin-Samuelson Theory of Trade." *Journal of International Economics* 7, no. 2 (May 1, 1977): 201–8. [https://doi.org/10.1016/0022-1996\(77\)90031-9](https://doi.org/10.1016/0022-1996(77)90031-9).
- . "The Non-Substitution Theorem and International Trade Theory." *AEPA Australian Economic Papers* 12, no. 21 (1973): 267–69.

REFERENCES

- Stokey, Nancy L. "Learning by Doing and the Introduction of New Goods." *Journal of Political Economy* 96, no. 4 (1988): 701–17.
- Strange, Susan. *International Economic Relations of the Western World. 1959-1971: International Monetary Relations 2*. Edited by Andrew Shonfield. London: Royal Institute of International Affairs, 1976.
- Swan, T. W. "Economic Control in a Dependent Economy." *Economic Record* 36, no. 73 (1960): 51–66.
- Tew, Brian. *The Evolution of the International Monetary System: 1945-81*. London: Hutchinson, 1988.
- Torrens, Robert. *An Essay on the External Corn Trade, Containing an Inquiry into the General Principles of That Important Branch of Traffic, an Examination of the Exceptions to Which These Principles Are Liable and a Comparative Statement of the Effects Which Restrictions on Importation and Free Intercourse Are Calculated to Produce Upon Subsistence, Agriculture, Commerce and Revenue*. London: J. Hatchard, 1815.
- Triffin, Robert. *Gold and the Dollar Crisis: The Future of Convertibility*. New Haven: Yale University Press, 1983.
- . *Gold and the Dollar Crisis: Yesterday and Tomorrow*. International Finance Section, Dept. of Economics, Princeton University, 1978.
- . *Tomorrow's Convertibility: Aims and Means of International Monetary Policy*. Rome: Banca nazionale del lavoro, 1959.
- Uekawa, Yasuo. "Some Theorems of Trade with Joint Production." *Journal of International Economics* 16, no. 3 (May 1, 1984): 319–33. [https://doi.org/10.1016/S0022-1996\(84\)80009-4](https://doi.org/10.1016/S0022-1996(84)80009-4).
- United Nations., and Michael Bruno. *Planning the External Sector: Techniques, Problems and Policies. Report on the First International Seminar on Development Planning, Ankara, Turkey 6-17 Sept. 1965*. New York: U.N., 1967.
- Viner, Jacob. *Studies in the Theory of International Trade*. London: George Allen & Unwin Ltd, 1960.
- Wade, Robert. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, New Jersey: Princeton University Press, 1990.

- Wilcox, Clair, and Havana United Nations Conference on Trade and Employment 1947-1948. *A Charter for World Trade*. New York: Macmillan Co., 1949.
- Williamson, John. "Surveys in Applied Economics: International Liquidity." *The Economic Journal* 83, no. 331 (1973): 685–746. <https://doi.org/10.2307/2230668>.
- Wolf, Martin. *India's Exports*. New York; London [etc.]: Published for the World Bank [by] Oxford University Press, 1982.
- World Bank. "Pakistan: Export Diversification and Trade Policy," 2012. http://VH7QX3XE2P.search.serialssolutions.com/?V=1.0&L=VH7QX3XE2P&S=AC_T_B&C=Pakistan%20:%20Export%20Diversification%20and%20Trade%20Policy&T=marc&tab=BOOKS.
- Young, Alwyn. "Growth Without Scale Effects." *Journal of Political Economy* 106, no. 1 (1998): 41–63. <https://doi.org/10.1086/250002>.
- Zaidi, S. Akbar. *Issues in Pakistan's Economy*. Oxford [u.a.: Oxford Univ. Press, 1999.

INDEX

A

Abramovitz, Moses, 295-96, 298-99, 315
Aghion, Philippe, 302, 315
Ahmad, Ehtisham, 249, 315
Albania, 193
Alexander, Sidney s, i, 3, 136, 169-70, 315
Alsace and Lorraine, 161
America, i, ii, 3, 5, 10, 35, 37, 95, 263-64, 266, 319
Amortisation, 225
Amsden, Alice H, 114, 116, 315
Arrow, Kenneth, 26, 32-33, 73-74, 286-87, 291-92, 300, 311, 315
Australia, 190
Austria, 95, 103, 161-64, 166, 193
Austro-Hungarian Empire, 161
Autarky, 86
Ayub Khan, Muhammad, President, 230, 235

B

Bacha, Edmar, 182, 315
Bagehot, Walter, 157
Balance of payments, 125
balance of trade, 125
Balassa, Bela, 181, 183-85, 315
Baluch, 268
Bangladesh, vi, 10, 224, 230, 235, 237, 265
Bank Act of 1844, 154
Bank for International Settlements, 222
Bank of England, 154, 156-57, 160
Banque de France, 157
Belgium, 159-62, 166, 189-90
Bello, Walden, 110, 117, 315
Bensusan-Butt, D. M, 11, 16, 49
Benthamite ideas, 268
Bhaduri, Amit, xv
Bhutto, Zulfikar Ali, President, 242
Bloomfield, Arthur I, 157

Bolshevik Revolution, 159
Bonus vouchers, 231, 235, 253
Branding, 20, 51, 55-57, 273
Brazil, 50, 67, 96
Brecher, Richard A, 40
Bretton Woods, 187
Bretton Woods system, xi, xii, 153, 187, 191-92, 199, 202-03, 205, 208, 217-18, 252
British Commonwealth, 189-90, 214
British India, 230
British pound, 201, 203
Brookings Institution, 215, 218
Brown, E. H. Phelps, 290-91, 315
Bruno, Michael, 184-85, 316, 326
Bullionist view, 154
Bundesbank, 222
Burmeister, Edwin, 284, 316

C

Canada, 42, 50, 190, 196
Cassel, Gustav, 159, 166
Castley, Robert, 113-14, 117, 263, 316
Census of Manufacturing Industries, 233
Central bank, 145, 148-52, 163, 165, 177, 188-89, 197, 203, 211, 213-14, 220-22, 239, 256-57, 259, 290
Chaebols, 115-16
Chain index, 277-78
Chairman of the Federal Reserve Board, 152, 221-22
Chamberlin, Edward, 273, 316
Champernowne, D. G, 277-78, 281, 284, 316
China, iii, vii, viii, 42, 65, 67, 102, 108, 117-19, 121, 131, 215, 219, 244, 260, 267
Chipman, John S, 26, 33, 38-40, 316
Choudhri, Ehsan U, 40, 258
Christian missionaries, 241, 245
Churchill, Winston Spencer, 157
Clark, J. B, 280
Clower, Robert, 197-98, 316

INDEX

CMI. See Census of Manufacturing Industries
Cohen, Avi J, 276, 316
Communism, 193, 268
comparative advantage, i, xiii, 1, 5-8, 26, 34, 49, 229, 235
 Definition, i
complex goods, viii, 67-68, 70-73, 75-77, 81-82, 84-86, 88, 95-96, 98-103, 105-06, 108-09, 111-14, 117-23, 229, 236-39, 242-45, 252, 263-65, 308
Complex goods
 Definition, 67
Consultative Group on Exchange Rates, 178-79
Consumer price index, 144, 257
Court of the Bank of England, 157
CPI. See Consumer price index
Cunliffe, Walter, Lord, 157, 159
Curcuro, Stephanie E, 217, 316
Currency School, 154
Cyhn, Jin W, 73, 115, 316

D

Dawes plan, 163
De Cecco, Marcello, 158, 317
de Gaulle, Charles, President of France, 213
De Grauwe, Paul, 158, 317
Debreu, Gerard, 32-33, 286
Defoe, Daniel, 3
Denison, Edward Fulton, 295, 298-300, 317, 320
Denmark, 58
Despres, Emil, 210
Differentiated goods, vii, ix, 49-51, 55-58, 60, 67, 140, 292-93, 301
Divisia index numbers, 297
Dobb, Maurice, 120, 317
Domestic resource costs, 184-85, 232-33, 315-16
Doner, Richard F, 115, 317
Dooley, Michael, 215-16
Dore, Ronald Philip, 108, 317
Dornbusch, Rudiger, 137-38, 317
DRC. See Domestic resource costs

Dudler, Hermann-Josef, 211, 213, 317
Dvorak, Tomas, 217, 316

E

East Pakistan, 230, 235-36
EBS. See Export Bonus Scheme
EC. See European Commission
ECB. See European Central Bank
Edgeworth, 2
Education, viii, xiii, 38, 42, 67, 70-71, 74-75, 100-03, 109, 118, 120, 127, 194, 229, 231, 238-39, 242, 244-45, 246-48, 259-60, 263, 266-67, 274, 287, 319, 322
 Training, 75
Effective protection, 183, 232
Effective rate of protection, 181-85, 232-33
Egypt, 237, 246, 264
Eichengreen, Barry, 161, 163, 166, 215, 317
Emmanuel, Arghiri, 46-48, 317-18
Endogenous growth, 76, 293, 301, 303-11, 313
Enos, John Lawrence, 114, 318
ERP. See Effective Rate of Protection
Ethier, Wilfred, 27, 29-32, 284, 318
Eurodollar, 200, 209-10, 214
Europe, xii, 3, 42, 59, 64, 84, 94-95, 105, 152, 156-158, 160-161, 164, 166-167, 187-189, 191-96, 199, 201, 205-06, 208, 210-14, 216, 221, 244, 246, 251, 263-66, 268, 321
European Central Bank, 221, 319
European Commission, 59, 63-64
European Common Market, 46, 97
European Economic Community
 EEC, 201
European Free Trade Area, 46, 97
European Payments Union, 193
European Recovery Program. See Marshall Plan
European Union, 63-64, 97, 237
Exchange Settlement Certificates, 230
Exorbitant privilege, 206, 213, 215-16, 318-19
Experience, 73

Learning by doing, 72
 Export Bonus Scheme, 230, 234-35,
 239-41, 253

F

Factor endowments theory, i, 1, 5-9,
 19, 21, 24-31, 33-34, 36-41, 43-44,
 46-50, 52, 56, 167, 172
 factor price equalisation, 27, 29, 34,
 41, 50
 FEBC. See Foreign exchange bearer
 certificate
 Federal Reserve, 149, 152-153, 157,
 160, 203, 213, 221, 222
 Felipe, Jesus, 280, 318
 Field, David, xv
 Financial Times, 118, 218
 Finland, 141
 First World War, xi, 35-36, 43, 94,
 154-56, 158, 167, 187, 192-93,
 220, 272
 Fischer, Stanley, 137-38, 318
 Fisher, Franklin M, 280, 318
 Fleming, Marcus J, 208, 318
 Fordney-McCumber Tariff, 166
 foreign direct investment, 65, 111,
 114, 118, 218
 Foreign direct investment, 65, 70, 72,
 98, 112, 121, 217-18, 244, 321
 Foreign exchange bearer certificates,
 256, 259
 Foreign remittances, 236, 267
 France, 35, 37, 49, 54, 156-57, 159-
 62, 166-67, 189-90, 193-94, 196-
 97, 200-01, 203, 206, 208, 318
 Frankel, Jeffrey A, 215
 French franc, 201
 Friedman, Milton, 172-78, 318

G

GAB. See General Arrangement to
 Borrow
 Garber, Peter, 215-16
 Gardner, Richard N, 192, 195, 318
 Garegnani, Pierangelo, 281, 318

GATT. See General Agreement on
 Tariffs and Trade
 GDP. See Gross domestic product
 GDP deflator, 144, 253-54, 256
 General Agreement on Tariffs and
 Trade, 45, 96-97, 100, 107, 195,
 242-243, 251
 General Arrangement to Borrow, 199
 General Services Tax, 249
 General System of Preferences (GSP), ii
 Generalised System of Preferences,
 46, 66, 237
 Genoa Conference, 160
 German chemical industry, 35, 36
 German reparations payment, 160
 Germany, ii, 36, 49, 54, 59-61, 94, 103,
 141, 152, 156, 160-66, 192-96, 200,
 209, 212, 220
 GFCF. See Gross fixed capital formation
 Ghana, iv, 224
 Ghaus-Pasha, Aisha, 248
 Giscard d'Estaing, Valéry, 208, 213,
 318
 gold exchange standard, xi, 202-03
 Gold exchange standard, 189
 gold standard, xi, 94, 153-54, 157-61,
 164, 166-67, 173, 187, 189, 191,
 200, 203, 206, 209, 269
 Goldberg, P. K, 59, 62-64, 318
 Gourinchas, Pierre-Olivier, 216-17, 318
 Government of Pakistan
 Economic Survey, 234, 236
 Great Depression, 94, 158, 272
 Greece, 193-94, 221-22, 264
 Greenspan, Alan, 152-53
 Gresser, Edward, 10, 319
 Griliches, Zvi, 295, 297-300, 317, 320
 Gross domestic product, 126-27
 Gross fixed capital formation, 127,
 239, 241
 Group of Ten, 206
 GSP. See Generalised System of
 Preferences
 Guisinger, Stephen E, 232, 322
 Gwadar, 242

INDEX

H

- Haberler, Gottfried von, 7-8, 43, 45, 172, 181, 235, 319
Habib, Maurizio Michael, 216-17, 319
Hahn, Frank H, 26, 32-33, 282-84, 286-87, 289, 311, 315, 319
Hamilton, Alexander, i, ii, 3-4, 8, 86
Haq, Mahbub ul, 231, 319
Haq, Zia ul, President, 242
Haque, Irfan ul, 43, 319
Harberger, Arnold C, 169-70, 319
Harcourt, G. C, 276, 316
Hard currencies, xii, 147-51, 174, 200, 210, 217-18
Harrod, Roy Forbes, Sir, 167-70, 296, 298, 319
Heckscher, Eli, i, iv, vi, vii, 1, 5-6, 19, 29-31, 33, 40, 167, 325
Heckscher-Ohlin theory. See Factor endowments theory
Helfferich, Karl, 163, 319
Helmers, F. Leslie C. H, 137-39, 317
Helpman, Elhanan, 50-54, 319
Holland, 49, 189-90, 315, 325
Netherlands, 60, 218
Homogeneous goods, ix, 29, 51-52, 54-55, 140-141, 292
Howitt, Peter, 302, 307-09, 315, 319
Hume, David, 143, 154
Hungary, 161-63
Hyperinflation, 152, 162-63, 167

I

- IBRD. See International Bank for Reconstruction and Development
ILO. See International Labor Office
IMF. See International Monetary Fund
Indeterminacy, 2, 274-75, 281, 285-86, 312
Index number problem
Measuring capital, 278, 280, 283
India, iii, vii, viii, xiii, 3, 42, 65, 67, 75-76, 86-87, 96, 102, 120-23, 190, 231, 235, 237-38, 243-44, 246-47, 267-68, 327

- infant industries, 3, 8-9, 88, 100, 121, 243
Inflation
Definition, 171
Intellectual property, viii, 55, 118, 292, 301, 311, 313
Interim Committee, 206
International Monetary Fund
Articles of Agreement, 189
International Bank for Reconstruction and Development, 191
International Clearing Union, 188
International Labor Office, 253-55
International Monetary Fund, 87, 99, 107, 123, 137, 144, 174, 178-79, 188-89, 191, 195-200, 204-06, 209, 217-18, 222, 224, 242, 247, 251-52, 256, 267, 315, 318, 320
Article IV, 206
International Trade Organisation, 195
Intertemporal general equilibrium, 33
Intertemporal general equilibrium, 32-33, 53, 125, 282, 286-87, 289, 311-12
Intra-European Payments Agreement, 190
Ireland, 35-38, 77, 218, 221-22
Irrigation, 9, 237, 240, 247, 308
Canal system, 237
Isard, Peter, 60-62, 320
Italy, 50, 64, 161, 193-94, 196, 264
ITO. See International Trade Organisation

J

- Japan, 42, 46, 49, 60, 65, 67, 95-96, 101, 103-09, 111-14, 116-20, 131, 141, 152, 191, 196, 212, 216, 220, 236, 244, 263-65, 316-17, 321, 323
Jersey, 218, 326
Johnson, Chalmers, 104-05, 320
Johnson, Harry G, 208
Jones, Charles I, 305-08, 320
Jorgenson, Dale W, 295, 297-300, 317, 320

K

Kahn, Richard Ferdinand, Lord, 208
 Kemp, Murray C, 31, 320
 Kenen, Peter B, 206-07, 215, 320
 Kennedy Round, 97
 Kennedy, John F, 202-03
 Keynes, John Maynard, 94, 150, 158,
 161, 164-65, 168, 188, 191-92, 205,
 276, 320, 324
 Khan, Azizur Rahman, 254
 Khan, Mohsin S, 258
 Kim, Linsu, 73, 115-17, 321
 Kindleberger, Charles Poor, 210-15, 321
 Knetter, Michael M, 60-62, 64, 321
 Komiya, Ryutaro, 105, 321
 Korea, ii, iii, vii, viii, 42, 65, 67, 101-
 02, 108, 112-15, 117-21, 131, 236,
 243-44, 260, 263-67, 316, 318
 Korean Boom, 231, 245
 Korean War, 112, 194
 Krause, Lawrence, 104, 321
 Krueger, Anne O, 184, 321
 Krugman, Paul R, 50-54, 59, 62, 177,
 178, 319, 321
 Kuwait, iv, 34, 219

L

Läidi, Zaki, 102, 321
 Lancaster, Kelvin J, 292, 304, 321
 Latin America, xi, 42, 97, 227, 264-65
 Law of one price, 53, 55, 76, 140, 141
 Leamer, Edward E, 39, 40, 321
 Learning by doing, 74, 291-92, 300,
 304
 Learning curves, 73
 Lend-Lease, 188
 Leontief, vi, 21, 24-25, 34, 38-40, 43,
 278, 321
 Levhari, David, 281, 322
 Lewis, Stephen R, 232, 322
 Lin, Ching-Yuan, 112, 230, 321-22
 linen industry, 36-38, 77
 Lipsey, Richard, 197-98, 316
 List, Friedrich, ii, 3-4, 8, 86, 322
 Lithuania, 164
 Local costs, viii, ix, 140

Long Term Arrangement, 97
 Low wage goods, 47, 68, 70-71, 81-82,
 84, 86, 88, 112, 229-30, 234,
 244-45, 260, 313
 Definition, 69
 Lynd, Douglas, 246, 322

M

Madagascar, 252
 Mahalonobis. P. C, 120
 Mainwaring, L, 28-29, 322
 Malaysia, iii, 65, 98, 236, 243, 260,
 264-265, 267
 Malta, 218
 Manchuria, 112
 Mangoldt, 2
 Maquiladoras, 98, 265
 Marshall Plan, 193
 Marshall, Alfred, 4, 5, 194, 272-73
 Marston, R, 62
 Martin, William McChesney, 149,
 221-22
 Mathur, Gautam, 120
 Mazzucato, Mariana, 310, 322
 McKinnon, Ronald I, 206, 213, 214-15,
 322
 Meade, James E, 44, 170-72, 181, 322
 Metcalfe, J. S, 28-29, 31, 47, 322, 325
 Mexico, 49, 76, 98, 227, 243-44, 265
 MFA. See Multi-Fibre Agreement
 MFN. See Most favoured nation
 Middle East, 127, 244, 259, 266
 Migration of workers, 263-64, 266
 Military expenditure, 194, 202, 212,
 214
 Mill, John Stuart, 2, 155, 156, 168,
 170, 308
 Miration of workers
 Scientists and engineers, 266
 Most favoured nation, 95
 Multi-Fibre Agreement, 66, 97, 108,
 234, 237

INDEX

N

Nasim, Anjum, 257, 322
Nazis, 95
New Zealand, 97, 190
Newton's laws of mechanics, 288
Nigeria, 246
Noble, Gregory W, 317
Non-government organisations, 261
Norway, 20, 141, 219

O

OECD. See Organisation for Economic
Cooperation and Development
OEM. See Original equipment
manufacturing
Ohlin, Bertil, i, iv, vi, vii, 1, 5-8, 19-
25, 29-31, 33-40, 42, 77, 164-65,
167, 320, 323, 325
Okuno, Masahiro, 105, 321
Organisation for Economic Cooperation
and Development, 107
Original equipment manufacturing, 90
Ozawa, T, 104, 107, 323

P

Pakhtoon, 268
Pakistan, 2, iii, vi, xii, xiii, xv, 65, 76,
97, 100, 105, 227, 229-39, 241-60,
263, 266-68, 315-16, 318-19, 321-
23, 325, 327
Papanek, Gustav F, 232, 323
Paris, 37, 156, 201, 318, 321, 323
Park, U-Hui, 114, 318
Pasha, Hafiz A, 248
Pasinetti, Luigi L, 272, 281, 295, 298,
310, 312, 323
Patents, vii, 55, 77-79, 81-82, 84, 86,
118, 122, 132, 292, 301, 311
Pearce, Ivor, 21
Peel, Robert, 154
Philippines, 237-238, 243-244, 246
Pigou A. C., 4, 158, 323
PL480, 112, 114
Polak, Jacques Jacobus, 204, 323
Poland, 95, 161, 163

Portugal, 1, 3, 4, 95, 189-190, 264
PPP. See Purchasing power parity
Pricing to market, 58-59, 62, 140-42
Production functions, xiv, xv, 5-7, 19-
20, 23-24, 28-29, 31, 33-34, 38-39,
44, 272-73, 276-77, 280-81, 283-85,
290-91, 293, 295-96, 298, 311-12
Cobb-Douglas, 290, 296, 298, 303,
315
Proprietary knowledge, vii, viii, 20,
35, 37, 76-79, 81-86, 88-93, 100,
103-17, 121-23, 239, 242-45, 265
Purchasing power parity, 159

R

Raj, K. N, 100-01, 120, 323
Ravenhill, John, 317
Read, L. M, 272, 298, 323
Real Equilibrium Exchange Rates, x
Real exchange rate, 137, 144, 172,
176-77, 179-80, 224
Equilibrium exchange rate, 172,
177-81, 224
Real equilibrium exchange rate,
172, 177-78
Reichsbank, 160
Rentenmark, 162
Reparations, 95, 161
Research, 107
Research and development, iii, vii,
viii, 36, 51, 68, 76-85, 89-93, 98,
104-06, 110, 115-18, 121-122, 136,
220, 272, 291, 294, 299-14, 325
Rey, H el ene, 216-217, 318
Rhee, Syngman, 108
Ricardian theory, i
Ricardo, David, i, iv, 1-8, 27, 50, 154-
156, 308
Rist, Charles, 150, 157, 159, 165, 323
Robinson, Joan Violet, 167-70, 273,
276-78, 323
Rodrik, Dani, 43-44, 324
Romer, Paul M, 301-03, 306, 308, 324
Rosenfeld, Stephanie, 110, 117, 315
Routh, Guy, 168, 324
Royalties, 104-05, 117, 132
Rueff, Jacques, 165, 324

Ruhr, 162
 Russia, 95, 120, 157, 159-61, 193
 Rybczynski theorem, 27, 30-31
 Rymes, Thomas K, 272, 298, 301, 324

S

Salant, Walter, 210
 Samuelson, Paul Anthony, 26-31, 41,
 44, 47-48, 280-81, 284, 320, 324-25
 Sargent, Thomas J, 163, 324
 Schimmelpfennig, Axel, 258
 Schumpeter, Josef Alois, 272, 311
 Schwartz, Jacob, 186, 289, 325
 Second World War, ii, xi, 42-43, 65-
 67, 73, 84, 94, 101, 103, 119, 153,
 165, 168, 187, 194, 222, 263
 Segerstrom, Paul S, 305, 325
 Seignorage, 213-15
 Sekiguchi, Sueo, 104, 321
 Self-reliance, 84
 Sen, Amurtya K, 100-01, 120, 323
 Senior, Nassau, 4
 Shadow exchange rates, x
 Short Term Cotton Textile
 Arrangement, 97
 Simple goods
 Definition, 67
 Sindh, 268
 Singapore, 42, 109, 219, 264
 Smith, Adam, i, ii, 1-4, 73, 291, 293,
 325
 Smithsonian "tunnel", 201
 Smithsonian agreement, 201
 Snake, 201
 Socialism, 268, 269
 Sohmen, Egon, 198
 Soligo, Ronald, 232, 325
 Solomon, Robert, 192, 202-03, 325
 Solow, Robert M, 278-84, 290, 295-98,
 300, 325
 De Vries lectures, 281, 290
 South Africa, 190
 Spain, 95, 189-90, 221-22, 264
 Special Drawing Rights, 204-06, 217
 Special Drawing Rights, xii, 174, 202
 Sraffa, Piero, 217, 271-76, 282-83,
 285-86, 289, 312, 325

Sri Lanka, 246
 State Bank, 256-57, 322
 Steedman, Ian, 28-29, 31, 47, 322, 325
 Stern, Joseph J, 232
 Stokey, Nancy L, 304, 326
 Stolper-Samuelson theorem, 27, 30,
 31, 44
 Strange, Susan, 212, 326
 Subcontracting, viii, 58, 83, 88-91,
 109-11, 113-15, 122, 131, 236, 265
 Surrogate capital, 280-81
 Surrogate production function, 281,
 284, 324
 Suzumura, Kotaro, 105, 321
 Swan, Trevor, 143, 280, 326
 Sweden, 5, 49
 Switzerland, 58, 106-07, 141, 218

T

Taira, K, 108
 Taiwan, ii, iii, vii, viii, 42, 65, 67, 97,
 101-02, 108-15, 117, 119-21, 220,
 230, 236, 243-44, 264, 322
 Taussig F. W., 4, 161
 Taylor, Lance, 182
 technical progress, iii, vii, 8-9, 20, 35-
 36, 63, 67-68, 74, 76, 85, 105, 117,
 178, 272, 277, 287, 291, 293-303,
 305, 306, 309-14
 Thailand, iii, 65, 98, 236, 243-45,
 264-65, 267
The Economist, 210
 Torrens, Robert, 1, 275, 326
 Total factor productivity, 297, 306
 Trademarks, 55, 78, 292, 301
 Triffin, Robert, 202, 204-06, 217, 326
 Tunisia, 65, 98, 260, 265
 Turkey, 65, 98, 249, 264-65, 326

U

UK
 United Kingdom, 49, 60-61, 94, 97,
 153, 187-188, 190, 194-197,
 199-201, 210, 218, 221, 230,
 233, 252
 Union of Soviet Socialist Republics, 94

INDEX

Soviet Union, 118, 120, 189, 204-05, 317
United Nations Education, Scientific and Cultural Organization, and United Nations Children's Fund (UNESCO), 246, 319
University of Copenhagen, 218
US
 United States, vi, xi, xii, xiii, 4, 6, 8, 20, 25, 34, 38-40, 42, 46, 49-50, 59-62, 64-65, 84, 94-95, 103, 105, 107, 109-10, 112-16, 119, 121-122, 131, 141, 149, 152-53, 156-61, 164, 166-67, 174, 187-96, 199, 201-203, 205-22, 231, 237, 242, 244-46, 250-51, 253-56, 262-66, 306, 308-09, 318
US Congress, 193, 195, 203

V

Vanek, Jaroslav, 25-26, 40, 51
Venice, 37, 49
Verboven, Frank, 59, 62-64, 318
VERs. See Voluntary export restraints
Versailles Treaty, 95, 162
Vietnam, 152, 245
Viner, Jacob, 4, 326
voluntary export restraints, 43, 45, 97, 105

W

Wade, Robert, 97, 109-11, 326

Walras, Leon, 5
Warnock, Francis E, 217, 316
West Pakistan, 230, 235
Western Europe, ii, 46, 95-96, 187, 191, 242
White, Harry Dexter, 94
Wholesale price index, 144
Wicksell, Knut, 155, 272, 275
Wicksteed, Philip Henry, 273
Wilcox, Clair, 190, 192, 195, 327
Williamson, John, 207, 327
Wisman, Jon, xv
Wolf, Martin, 86, 327
World Bank, xv, 87, 99, 107, 123, 191, 193-195, 227, 236-237, 240-42, 250-51, 267, 317, 327
World Trade Organisation, 87-88, 93, 95-96, 100, 196, 242-243, 251
Worswick, George David Norman, 282
WPI. See Wholesale price index
WTO. See World Trade Organisation

Y

Yang, Jiawan, 62, 321
Young, Alwyn, 306-08, 327
Yugoslavia, 193


Z

Zaidi, Akbar, 241, 327
Zollverein, ii
Zolnierenek, James M, 305, 325

ERRATUM

The restrictions on travel and access to libraries prevented the author from completing all the references in the text. Because of that the following reference, pertaining to pages 41–42 of the text, could not be added in time for publication:

Becker, Gary. Human Capital and the Personal Distribution of Income. An Analytical Approach. W. S. Woytinsky Lecture No.1. University of Michigan.



Some economic facts are so familiar that one usually forgets to ask why they should be and, yet, mainstream theory cannot explain them. Why are nearly all tradable capital goods and complex manufactures produced by firms of the advanced economies of the rich countries and East Asia? Why can these economies have far greater international trade and finance imbalances with flexible exchange rates than they could with fixed exchange rates? Do comparative advantages explain the successes of Japan, South Korea and China? Why can poor countries never extricate themselves from their foreign debts, despite giving priority to exporting? Why have their currencies been continually depreciating with no lasting improvement of their exports or foreign debts? And so on.

This book shows how the questions can be convincingly answered with simple, rigorous, step by step reasoning starting from the facts that wage rates differ widely, most technical progress is from R&D, especially of firms, and capital goods are traded. Mainstream theory persists despite widespread dissatisfaction with it because it purports to provide a logically coherent framework. This book refutes the "logic" by meeting the theory on its own ground and laying bare the fallacies. And the description of the consequences of applying the theory in Pakistan will look familiar to many poor countries.

As rich countries form divisive trade blocs, ecological degradation spreads, income inequalities increase and economies succumb to pandemics it is time to scrap the fantasies of equilibrium and efficient resource allocation of mainstream theory and reflect anew.



PKR 2000

ISBN: 978-969-7502-08-0